

Lessons from an epidemic

Britain's outbreak of foot-and-mouth disease has revealed failures by senior civil servants to learn from previous experience. It has also opened up an unfortunate rift between epidemiologists and veterinary scientists.

As the British outbreak of foot-and-mouth disease begins to subside, recriminations are flying. More than 3.4 million animals have been slaughtered; the cost to farming and tourism runs to billions of pounds. Opposition politicians are calling for a public inquiry. But the government seems likely instead to launch a more limited scientific review of how to prevent animal disease, probably entrusted to the Royal Society.

Most of the current debate surrounds measures used to control the outbreak. The 1969 Northumberland Report into Britain's last foot-and-mouth crisis in 1967 concluded that restrictions should be imposed on animal movement as soon as possible, and that the military be deployed if any logistical problems arise in the culling and disposal of animals. Any investigation should ask why it took the Ministry of Agriculture, Fisheries and Food (MAFF) three days to impose movement restrictions after the first case was confirmed. It should also determine why the army did not become involved until almost a month later.

MAFF released epidemiological data in weeks (rather than years, as happened in the bovine spongiform encephalopathy epidemic). But the largest share of credit for this goes not to MAFF, but to John Krebs, who heads the Food Standards Agency. An animal disease of primarily economic significance is, strictly speaking, beyond the agency's remit, which is to protect public health. But Krebs called together leading scientists to discuss the disease within two weeks of the first confirmed case. Data were then requested from MAFF, and placed in the hands of the epidemiologists convened by Krebs. These scientists' subsequent work was crucial in revealing that a more rigorous culling policy was

needed to control the epidemic (see *Nature* **410**, 515–516; 2001). They went on to work under David King, the government's chief scientific adviser, who assisted ministers in their policy options. Although MAFF has now been abolished (see *Nature* **411**, 727; 2001), its officials permeate the new Department for Environment, Food and Rural Affairs. They should abandon their territorial leanings and become more proactive in seeking appropriate scientific help.

The veterinary community has also revealed its territoriality. British veterinary scientists — both practitioners and researchers — seem resentful of the leading role played by epidemiologists in the current crisis. Today, some are questioning whether so many animals needed to be killed, others are asking whether vaccination might have been used, and yet others are arguing that the epidemiologists' models were insufficiently sensitive to the virus's behaviour in different host species, and so on.

In few cases do those asking such questions have a coherent argument that alternative policies would have produced a better outcome — and the suspicion is that the complaints are motivated by a desire not to let scientific 'outsiders' take credit for delivering generally sound advice.

That is unfortunate, because veterinary scientists must now work with the wider scientific community to develop the tools required to fight future outbreaks — in Britain, and elsewhere. Most urgently needed are diagnostic tests that can quickly identify animals incubating the disease, and comprehensive databases containing detailed information on animal holdings, demographics and movements. ■

Defending tax-funded navigation

A congressional committee has erred in its appropriate desire to support free enterprise.

Just as debate is raging in the life sciences over models for enhanced access to the full text of the scientific literature, a budget cut recommended by a US congressional subcommittee on energy and water development, of all things, threatens the most basic of researchers' services: search functions across authors, titles and abstracts. The committee singled out PubScience, developed by the Department of Energy to provide such services across the physical sciences and modelled on the biomedical service PubMed. The committee decried PubScience as an undesirable duplication of activities already carried out by the private sector. If enacted, its recommendation will mark the death knell of PubScience.

PubScience, barely two years old, has yet to establish itself. There is little doubt that political lobbying by large secondary publishers (see page 980) influenced the recommendation, and that this exercise has been a practice run for a subsequent challenge on the more established PubMed. Given PubMed's strong bipartisan support across US politics, a challenge seems unlikely to succeed, and seems rather to be aimed at preventing any further expansion of PubMed.

Proponents of a strong government role in scientific information and in a free-access archive to the entire literature should take note of congressional opposition to their position. Supporters and benefi-

ciaries of a free and competitive market, including this journal, might be tempted to salute the congressional recommendation, which upholds a principle that efficiency and public interests are generally best served by governments outsourcing to a competitive private sector, rather than trying to emulate it.

But PubMed, run by the National Library of Medicine, can be appreciated by anyone in biology as a service that works, not only for researchers but also for the public. Publishers who contribute abstracts to both PubMed and PubScience do so voluntarily, and both services drive traffic to publishers' subscription sites. These services provide no-frills access across the literature that, if left to the private sector, would have been obtained more slowly and at a greater cost to the research enterprise.

Those dismayed by the latest development would be wrong simply to blame the publishers, who have as much right to lobby Congress as high-profile researchers. More to the point, a resistance to government incursions on free enterprise is common ground between Democrats and Republicans. But the real issue is where lines should be drawn, and this requires a subtle judgement of the public interest, which is why a bill being introduced that would support PubScience has a chance of success. ■





Under threat
Budget proposal
could spell the end
for PubScience
p980



Just say yes?
Concern increases
over bias in drug
approval process
p981



Price war
Developing world
takes a stand on
cheap medicines
p982



Cotton crisis
Transgenic cotton
rejected by Indian
government
p983

Physicist nominated as science adviser for US administration

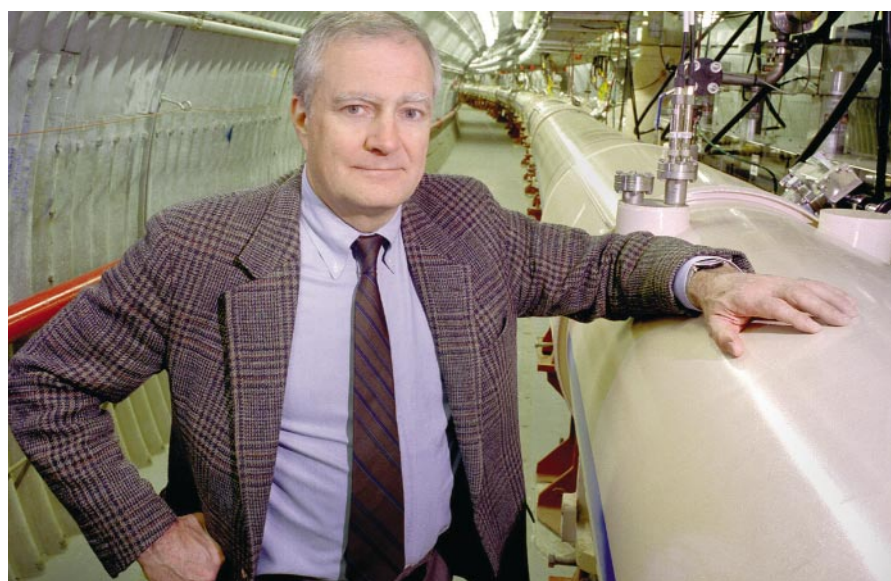
Tony Reichhardt, Washington

President George W. Bush ended months of speculation on 25 June, when he nominated John Marburger, a physicist who now directs the Brookhaven National Laboratory on Long Island, New York, for the post of White House science adviser.

Pending approval by the US Senate, 60-year-old Marburger, whose research background is in lasers and nonlinear optics, will also direct the White House Office of Science and Technology Policy.

In an interview, Marburger told *Nature* that he sees the adviser's role as being that of "a broker between the scientific community and the administration". But at the same time, he cautions that "science is not the only driver for policy", and that the president will always consider other factors, such as economics, in his decisions. And, ominously for scientists looking for extensive budget increases, Marburger notes that Bush's recently enacted tax cut could make such increases a difficult proposition.

But with the White House already ensnared in scientific controversies ranging from climate change to stem-cell research (see below), it is Marburger's established



Words of advice: John Marburger says scientists need to become more responsive communicators.

skills as a mediator and public explainer of science that are likely to come to the fore.

Marburger believes that scientists need to be better communicators — and to respect others' points of view. "It isn't as if scientists

[just] need to teach the public about science," he says. "There's a lot of emotion on these issues, and you need to be patient with the concerns and take them very seriously. You can't just stonewall people who are

Stem-cell debate heats up as Bush ponders policy

Meredith Wadman, Washington

With the Bush administration poised to decide in July whether to allow federal funding for research on human embryonic stem cells, leading scientists have stepped up criticism of the idea that adult stem cells could offer an alternative.

At a National Academy of Sciences workshop on 22 June, researchers questioned the strength of recent research that suggested adult stem cells can switch function. The critics included authors of papers that appeared to show that stem cells from adult tissue in mice and humans can transform into fully differentiated cells of other tissues.

Speakers at the workshop said that the existence of such transformations remains

unproven, and efforts to repeat some of the experiments have failed. In mice, on which many of the experiments rely, "all of the early claims are now falling apart", said Irving Weissman, a biologist at Stanford University.

Margaret Goodell, the senior author on a 1999 *Proceedings of the National Academy of Sciences* paper that helped to launch the early excitement about adult stem cells, said that she felt her work was being misused for political purposes. "The science doesn't justify what's happening at the moment, which is [people] saying 'adult stem cells can do anything'," said Goodell, a stem-cell biologist at Baylor College of Medicine in Houston.

But opponents of embryonic stem-cell research noted that the scientific community

does not speak for the moral status of human embryos. Kevin Fitzgerald, a molecular geneticist at Loyola University Medical Center in Chicago, said that it was not necessarily right for scientists to pursue all available lines of research, and called for embryonic stem-cell research to be confined to animals.

President Bush last week publicly backed a bill before Congress that would criminalize cloning for research or reproductive purposes (see *Nature* 411, 3; 2001).

But senior Republicans who oppose abortion are urging Bush to fund embryonic stem-cell research. Senate minority leader Trent Lott (Republican, Mississippi) said on 24 June that its benefits were "substantial" and "should be carefully considered". ■

concerned about their health, or who are concerned about their faith.”

Marburger arrived at Brookhaven in 1998 in the role of clean-up man, following a huge public row over a radioactive leak that had mobilized local opinion against the laboratory (see *Nature* **400**, 303; 1999). He soon received widespread praise for calming down what had become an ugly feud between laboratory staff and environmental activists.

Before Marburger took over, “you couldn’t even get people at the lab to return your phone calls”, says Scott Cullen, a lawyer for the anti-nuclear STAR Foundation, which led local criticism of the laboratory. “It was a completely adversarial relationship,” Cullen adds. “He definitely changed that.”

It wasn’t Marburger’s first experience in a hothouse public controversy: 20 years ago, he chaired a fact-finding panel on Long Island’s Shoreham Nuclear Power Plant, which was bitterly opposed by activists and was eventually dismantled.

Although lacking direct authority over major research agencies, such as the National Science Foundation and the National Institutes of Health, the science adviser is the US administration’s top-ranking official for setting science policy. If, as expected, Marburger is confirmed by the Senate, he will also co-chair the President’s Committee of Advisors on Science and Technology with the venture capitalist Floyd Kvamme (see *Nature* **410**, 617; 2001).

Marburger received his undergraduate degree from Princeton and his doctorate in applied physics from Stanford. In the 1970s he was a department chair and dean at the University of Southern California, where he co-founded a centre for laser studies. From 1980 to 1994, he was president of the State University of New York at Stony Brook, from where he also served as chairman of the Universities Research Association as it managed construction of the ill-fated Superconducting Supercollider project.

Michael Lubell, head of public affairs at the American Physical Society, calls the nominee “a good choice”, and praises him for his solid understanding of science policy. But Lubell worries whether the new science adviser will obtain direct access to the president, or instead will have to work through White House staff director Andrew Card.

It is a question Marburger has asked himself. “I was interested in how the staff works to provide advice to the administration on all kinds of issues,” he says. “And I was impressed with the dynamic that I found in the president’s office. They work well together as a team, and the president seems to listen.”

Budget proposal casts doubt over physics portal’s future

Declan Butler

A powerful congressional committee has passed a budget bill which, if enacted, could close down PubScience, a free search service for the physical sciences literature, operated by the US Department of Energy (DoE).

The budget bill, which was passed last week by the House appropriations subcommittee for energy and water development and was expected to be endorsed by the full appropriations committee, is likely to have a chilling effect on other government-operated services, including the National Library of Medicine’s PubMed Central, according to some observers.

The DoE’s Office of Scientific and Technical Information (OSTI) launched PubScience in 1999 to give physical scientists the kind of free, journal cross-searching facilities already offered to life scientists by PubMed. But the House subcommittee says in its bill that it is concerned that the OSTI is duplicating information services already available from commercial publishers, and urges a careful review of PubScience. This advice is likely to be heeded: “Needless to say, OSTI and DoE will be very responsive to the guidance of Congress,” says Walter Warnick, director of the OSTI.

The subcommittee proposes a 2002 budget for the OSTI of \$7.9 million, \$1.1 million less than it asked for, and more than \$700,000 less than this year’s budget. This cut is greater than the running costs of PubScience. Before becoming law, the entire bill has to be approved by the full House, agreed with the Senate and signed by President George W. Bush.

The proposal arose after a lobbying campaign aimed against PubScience, spearheaded by the Washington-based Software & Information Industry Association (SIIA) on behalf of for-profit and non-profit member companies including Reed Elsevier, ISI, Chemical Abstracts Services and Cambridge Scientific Abstracts.

David LeDuc, an SIIA official, says that PubScience provides a service similar to products offered by the association’s members, and “makes it increasingly difficult for these private-sector companies to continue offering their products”. He says that government initiatives should confine themselves to providing access to government information, and not act as secondary publishers.

Arie Jongeman, managing director of Elsevier Science’s physical sciences division, says that the subcommittee’s proposal means that “people have come to their senses; sanity is prevailing”. He claims that services such as PubScience are often less efficient than outsourcing, and that their true costs can be



Open to suggestions: Walter Warnick says he will take the congressional advice seriously.

concealed. Jongeman adds that the appropriators’ action “is an important signal that this kind of US governmental support is something which is dangerous and can flip back to zero overnight”.

But one DoE official argues that PubScience is “well anchored in law”, noting that the department’s mission explicitly includes the dissemination of scientific information.

And Martin Blume, editor-in-chief of the American Physical Society, which has extensive publishing interests, says that PubScience is not in competition with these “any more than the National Library of Medicine, MedLine, and PubMed are”. He adds: “The concern of the appropriators is not justified.”

But the SIIA is widely expected to attack PubMed next. The better-established life-sciences service, however, may prove to be a difficult target, as the National Library of Medicine, which operates it, has very strong support in the Congress.

There are signs, meanwhile, that the government’s role in scientific publishing is becoming a partisan issue. In May, Joe Lieberman (Democrat, Connecticut), chair of the Senate Committee on Governmental Affairs, introduced a bill that would give federal research agencies an explicit mandate to carry out PubScience-like activities. The act explicitly states that agencies should develop websites with links to the servers of outside publishers. The contrast between the two measures would appear to set the stage for a conflict on Capitol Hill between Democrats, who control the Senate, and Republicans, who control the House.

♦ <http://pubsci.osti.gov>

♦ <http://www.siiia.net>

♦ <http://www.nature.com/nature/debates/e-access>

Fear of bias puts spotlight on drug approval

Meredith Wadman, Washington

The past decade has seen the pace of drug approval by the US Food and Drug Administration (FDA) accelerate dramatically. But the agency is now facing mounting scrutiny over possible conflicts of interest in its approval process.

In preparation for a hearing later this year, staff investigators on a powerful congressional committee have gathered thousands of confidential FDA documents detailing the financial interests of the doctors and scientists who advise the agency on the approval of new therapies.

"We have to be absolutely sure that there are no conflicts of interest in the drug-approval process," says congressman Dan Burton (Republican, Indiana), whose House Committee on Government Reform is planning the hearing. "If the American people thought that there was even a possibility that the drugs approved by the FDA made it to market because some doctor or scientist had a financial stake in the products, then they would lose confidence in the entire system."

Separately, an activist group is preparing to sue the agency to force it to divulge more information on the conflicts of interest notified to it by members of its advisory panels.

Both the pharmaceutical industry and the FDA say that current law, which requires the FDA to screen advisers for financial conflicts, adequately protects the drug-approval process from bias. "We are certainly open to remedial recommendations. But first there has got to be proof that there's a problem. And we haven't seen that," says Jeff Trehwhitt, a spokesman for the Pharmaceutical Research and Manufacturers of America.

Because there is a shortage of experts who lack any vested interest in the pharmaceutical industry, the FDA often, and legally, grants 'waivers'. These allow experts who have financial interests in a company with a drug under review — or one of its competitors — to sit in judgment on the drug.

A study published last summer by the newspaper *USA Today* looked at conflicts of interest among some 300 experts on the 18 FDA committees that advise on drug approvals. It found that during two-and-a-half years beginning in January 1998, the agency waived conflict-of-interest rules 800 times. At 102 meetings dealing with the fates of specific drugs, 33% of the voting members had admitted a financial stake in the outcome.

Linda Ann Sherman, director of advisory-committee oversight at the FDA, says the study was flawed because it did not take account of confounding variables. For

instance, she says, most waivers are for advisers' interests in competing firms, not the firm whose drug is under consideration. And committee members have to declare as conflicts any affiliations their employers may have with drug companies. Advisers' financial conflicts, she says, "are for the most part indirect".

The existence of a conflict, and the fact that a waiver has been granted, is announced publicly at the approval meetings, but the nature of the conflict is rarely disclosed and neither is the amount of money involved.

Public Citizen's Health Research Group, meanwhile, is preparing to sue the FDA to force it to make this information public. The FDA says that committee members are exempt from having personal financial information disclosed under the Freedom of Information Act. "It should all be public," says Sidney Wolfe, director of the health group. ■



On trial: are financial conflicts of interest undermining the approval process for new drugs?

TEK IMAGE/SPL

US army advised to soldier on with biotechnology

Corie Lok, Washington

The soldier of the future is likely to be equipped with sensors embedded in the body to detect biochemical changes, as well as the means to automatically administer a customized drug cocktail in response, biologists have told the US army.

But to realize this vision, the army needs to build up far stronger expertise in molecular biology than it has at the moment, says a report published on 20 June by the National Research Council (NRC).

"The army needs in-house people to keep on top of how biotechnology is developing and how it impacts on what the army is trying to do," says Michael Ladisch, professor of agriculture and biological engineering at Purdue University in Indiana, who chaired the committee that produced the report.

According to the NRC panel, the army could use biotechnology to transform its fighting capability over the next 25 years. In that time, the experts believe that troops could be carrying sensors to detect chemical or biological agents — or even molecules signalling enemy presence — and wearing light armour derived from organic materials.

The report, *Opportunities in Biotechnology for Future Army Applications*, says that high-priority research areas for the army should include self-replicating biomaterials for wound healing, protein-based electronic devices for portable data storage, and drugs to treat shock. Areas of medium priority include edible vaccines and efficient solar cells based on photosynthesis.

The army should also support basic research in areas with military potential —

such as proteins for use in miniaturized, high-speed, radiation-resistant electronics — that are unlikely to be studied by biomedical researchers, the report adds.

The US army currently maintains extensive clinical research programmes but has had limited interest in basic molecular biology. The Department of Defense, of which the army is part, relies on the Defense Advanced Research Projects Agency (DARPA), with an annual budget of \$2 billion, to support science that may ultimately benefit the armed forces. But according to Verne Lynn, a member of the NRC panel and former director of DARPA, most of its biology research relates to just one function — protecting personnel against biological weapons. ■

► <http://www.napa.edu/books/0309075556/html>

Europe implements package deals for space station

Sally Goodman, Paris

Industrial researchers wanting to use the International Space Station (ISS) are being offered cut-price passage by the European Space Agency (ESA). The agency is offering rack space attached to the craft at prices from US\$60,000 per day.

ESA officials think they can attract commercial customers — who might want to use the gravity-free station for anything from materials research to drug development — by offering a more accessible and tailored service than NASA.

NASA only offers full equipment racks containing about one cubic metre of space in its Destiny laboratory for \$20.8 million for a full year. But at the ISS Forum 2001 in Berlin earlier this month, ESA published its price list, offering smaller lockers on its Columbus laboratory for as little as three months, with rates starting at \$715,000 for that period. External racks can be used for even shorter periods, at \$60,000 per day. The facilities should be available from 2005.

Jochen Graf, head of the ESA department responsible for selling commercial access, insists that the agency is not competing with NASA. "We are trying to expand the market and to move away from the classic space-science and industry customers," he says. European customers want smaller racks and time slots, Graf adds. "Three months is a good time period for customers using biological material who may want to bring down fresh material at regular intervals," he says.

ESA and NASA have each allocated 30% of their laboratory space to commercial customers. The European agency recently signed a \$1.7 million, four-year contract with Intospace, a company based at Hannover in Germany, to market the station to commercial users.

Graf said that, although European companies would be its main target, ESA was open to proposals from all over the world, including offers from public-private partnerships. He added that a request could be passed on to one of the other ISS partner agencies if ESA could not handle it.

The Russian Space Agency also plans to provide a competitive commercial package on the station. Russia expects to be able to take experiments there more cheaply using its Soyuz or Proton rockets than either NASA or ESA will be able to do on the US space shuttle. **S.G.**

♦ <http://www.esa.int/spaceflight/isscommercialisation>



Crisis point: countries facing killer diseases such as malaria want freedom from drug patent rules.

Poor nations push for right to produce cheap medicines

Sally Goodman, Paris

The United States clashed with developing countries over the impact of global trading rules on the distribution of medicines, at a special meeting of the World Trade Organization (WTO) in Geneva, Switzerland.

The meeting on 20 June, called by the WTO's African members, looked at whether the agreement known as TRIPS (trade-related aspects of intellectual property rights) should be modified to allow poorer countries better access to affordable drugs. TRIPS is aimed at protecting intellectual property rights internationally.

Officials said afterwards that, despite broad differences on the issue between developing countries and the United States in particular, further discussions would be held in the hope that global trade ministers can produce a clear statement on the issue when they meet in Qatar this November.

Pharmaceutical companies have argued, for example, that the TRIPS agreement entitles the United States to charge Brazil with breaching WTO rules by manufacturing its own generic AIDS drugs. The United States has so far refrained from taking such action.

According to the current TRIPS agreement, with which all WTO members must comply by 2006, compulsory licensing — the production of cheaper, generic versions of a drug — can take place without the prior consent of the patent holder only "in national emergencies or other circumstances of extreme urgency".

The developing countries want the WTO to state clearly that this allows them to act freely in responding to AIDS and other health crises. Otherwise, they say, TRIPS may have to be changed. They also want more time to comply with the agreement.

Delegates from European Union countries also urged clarification on some points

of TRIPS. They called for more discussion to produce a consensus on how it should be interpreted.

The US delegation said that it "would not object" if WTO member countries used the flexibility within the agreement to tackle major health crises. But it warned against the use of practices such as compulsory licensing "in the normal course of doing business". The delegation denied that the TRIPS agreement is not sufficiently clear or flexible.

David Earnshaw of Oxfam International says the outcome of the meeting was "entirely satisfactory" and has "shown up the United States as completely isolated in its position". Francisco Cannabrava, the Brazilian delegate, although critical of what he regards as the United States' inflexible reading of the TRIPS agreement, welcomed the fact that they were participating in the discussion.

But Harvey Bale, director-general of the International Federation of Pharmaceutical Manufacturers Associations, warned that developing countries, and not the pharmaceutical companies, could be the biggest losers if changes were made to TRIPS. "The world needs to decide where it wants pharmaceutical companies to put their R&D," he says. "There are lots of diseases with currently ineffective treatments where they could put their money instead."

Many countries and non-governmental organizations now hope that a declaration will be agreed in Qatar stating that the TRIPS agreement should not stand in the way of public-health agendas. They say this will help to stave off growing public criticism that the WTO is indifferent to the needs of poor countries. As a result of last week's meeting, two further WTO meetings on access to medicines have already been planned for between now and November. ■

♦ <http://www.wto.org>

India blocks sale of transgenic cotton seeds

K. S. Jayaraman, New Delhi

India has halted the commercial release of a genetically modified (GM) variety of cotton just two months after it became the first transgenic crop to be approved as safe by government scientists.

In a potentially serious setback for the global acceptance of GM crops, India's environment department declined to approve the commercial use of a cotton variety developed by Monsanto and its Indian subsidiary, Maharashtra Hybrid Seed Company (Mahyco). The department said it was not fully convinced of the biological safety or economic benefits of the variety, and has ordered fresh trials.

The decision is a snub for India's Department of Biotechnology (DBT), which had earlier said the cotton was safe and had recommended large-scale cultivation.

The cotton variety in question carries Monsanto's proprietary gene from *Bacillus thuringiensis* (Bt), which confers protection against the deadly bollworm (*Helicoverpa*



Nipped in the bud: India's approvals committee wants cotton trials to be conducted on a larger scale.

armigera). Almost 40% of pesticide use in India is aimed at controlling this single pest. India accounts for more than 15% of the world's cotton supply, and Monsanto says the use of Bt cotton will increase production and cut the use of chemical pesticides.

Since 1998, Mahyco has conducted field trials of the cotton in 52 locations, under DBT supervision. After getting the go-ahead from the department, Mahyco sought the required approval from the environment ministry's Genetic Engineering Approval Committee (GEAC) for commercial cultivation.

But the committee said on 19 June that the trial data were unreliable, as the crops were grown off-season when the pest load was naturally low. It asked the company to repeat the field trials on a larger scale — this time under the supervision of the Indian Council of Agricultural Research (ICAR).

Although the decision is a blow to Mahyco, which has spent US\$8 million over 6 years to develop the cotton, managing director Raju Barwale says that the company "looks forward to continue working closely with all government agencies and the ICAR in order to make this technology available to Indian farmers as soon as possible".

Sources close to the committee say that concerns about managing bollworm resistance raised by Greenpeace International science adviser Doreen Stabinsky played a part in its decision. They add that officials from India's health ministry also had reservations about the proposed use of a streptomycin-resistance gene as a marker, as this could potentially render the antibiotic useless.

The GEAC ruling has been welcomed by environmental groups such as the Delhi-based Research Foundation for Science, Technology and Ecology, which said in a statement: "Giving clearance would have opened the floodgates of other GM seeds and crops, when India does not have the appropriate technical and scientific infrastructure to deal with the risks of genetic engineering."

Kuldip Chopra, research director of Mahindra, another company that is developing herbicide-resistant corn, said: "We do not know the full facts but we had expected that Mahyco's cotton would at least receive permission for limited commercialization."

Curbs at NIH frustrate travel plans

Corie Lok, Washington

Thousands of researchers working for the intramural programme of the National Institutes of Health (NIH) could have their travel plans curtailed under reporting rules imposed by the US health department.

Travel to meetings that are to be attended by five or more NIH intramural employees must now be approved in advance by the office of Tommy Thompson, secretary of the health department of which the NIH is part. Any trip costing more than \$2,500 must be similarly approved. And travel to international meetings on AIDS and other infectious diseases, as well as visits to certain countries, including China and South Africa, also require central approval.

The rules on international travel were quietly implemented by the department in March and were extended last month to include domestic travel. Previously, NIH's postdoctoral employees needed only to clear their travel plans with their own institute.

Senior NIH officials say the rules simply reflect the new administration's attempts to gather information on travel. But scientific societies fear they will prevent researchers from attending conferences and doing their work effectively.

"There's concern both within NIH and outside that these rules will have a negative impact on the ability of scientists to freely exchange ideas with their colleagues in the government," says Martin Frank, executive director of the American Physiological Society. Frank says he has received complaints from NIH scientists and that his society plans to send the health department a letter about its concerns.

"It's hard keeping up with scientific research anyway," says Tony Mazzaschi of the Association of American Medical Colleges. "To add another hurdle is really untenable. I hope the rules are short-lived."

Campbell Gardett, a spokesman for the health department, said the rules were drawn up to look at questions about travel to glamorous places and the need for several government employees to attend the same meeting. "Travel should not be frivolous," says Gardett. "It should be necessary."



Vandals renew UK campaign against transgenic crops

London Vandals campaigning against genetically modified (GM) crops have destroyed plants at a British crop research centre, prompting fears that trials of the crops may face another summer of attacks.

Protesters destroyed a plot of GM barley at the John Innes Centre, Norwich, over the weekend of 15–17 June. Chris Lamb, director of the centre, says the plants were experimental research material and never intended for commercialization.

Pharmaceutical company Aventis has reported that several of its trials of GM oil seed rape have also been damaged in recent weeks. Other protesters trampled a huge cross in a GM crop trial site in the Black Isle in the north of Scotland on 7 June, the date of the UK general election.

More interest in AIDS drug access plan

New York Momentum seems to be building for a collaboration between the Joint United Nations Programme on HIV/AIDS (UNAIDS) and five drugs companies that

would give developing countries access to low-price AIDS drugs.

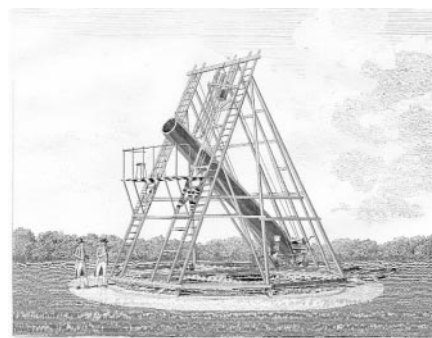
So far, 58 countries have expressed an interest in the project, this week's special UN session on AIDS, held in New York, was told.

The scheme aims to allow poorer nations to buy the drugs at reduced prices while protecting the intellectual property rights of the drug companies involved (see *Nature* 405, 263; 2000). Eleven countries have reached agreement with the companies and are receiving the drugs, UNAIDS says.

These countries do not include India or Brazil, both of whom already produce local versions of the drugs. Most international pharmaceutical companies are keen to limit the extent to which other countries import these lower-cost medicines, a process known as parallel importing.

Herschel telescope heads to the States

Washington One of history's most famous telescopes arrives this week at the Smithsonian Institution's National Air and Space Museum in Washington DC. A 20-foot reflector telescope built by the German-born astronomer Sir William Herschel in 1783 will be the centre-piece of a new exhibition, *Explore the Universe*, which will explain the history of cosmology.



View of the past: Herschel's telescope is set to play a starring role in the Smithsonian's exhibition.

On loan from the National Maritime Museum in Greenwich, London, the wooden telescope with a metal mirror was used by Herschel to catalogue nebulae and star clusters. The exhibition opens in September.

Victoria offers to pay for Australian synchrotron

Sydney The state of Victoria has stolen a march on its rivals in the race to host Australia's new synchrotron, after making a surprise offer to fund the project without federal support.

Half of the funding for the new A\$160 million (US\$83 million) synchrotron, Australia's largest science project for decades,

was set to come from the federal government. Queensland and New South Wales, along with the Australian Nuclear Science and Technology Organisation, Australia's national nuclear-research body, have already offered to raise the remaining funds and to host the facility.

But Victoria's premier, Steve Bracks, pre-empted a forthcoming announcement on who had been chosen by offering to host the project using a combination of state funds and private investment. Queensland's premier, Peter Beattie, has since called for a meeting of all the bidders in order to "get the process back on a proper footing".

Bush attacks genetic discrimination

Washington US President George W. Bush has announced plans to ban genetic discrimination by employers and insurers.

In his national weekly radio address on 23 June, Bush said that employers and insurers should not be allowed to use a healthy person's genetic profile, which may indicate a predisposition for certain diseases, to deny that person a job or insurance coverage, or to charge excessive premiums. Bush's announcement will encourage supporters of two bills currently being considered by the House of Representatives and the

Senate, both of which aim to ban genetic discrimination.

A 1996 law provides limited protection against this, and some states have their own restrictions on the use of genetic tests. But proponents of the bills say that comprehensive federal legislation is needed to fill in the gaps in state and federal law.

Astronomers see red over bright idea to tackle crime

San Diego Astronomers in southern California are marshalling forces to fight plans to install new street lights that would limit observations from local telescopes.

A San Diego city council committee last week approved a surprise measure calling for the phasing out of yellow sodium-vapour street lights, which minimize light beamed upwards. The committee claims that brighter lights are needed to help tackle crime in the city. Researchers at the Palomar and Mount Laguna observatories, both of which are within 100 km of San Diego, say the new high-intensity lights will limit the effectiveness of their telescopes. The council will make a final decision later this summer.

Astronomers fear that if San Diego does return to using the brighter street lights, governmental jurisdictions in southern California may follow.

Mice score instant tan with diet supplement

Washington Genetically engineered mice that change colour at the flick of a dietary switch have been created by researchers at the University of Virginia in Charlottesville.

Heidi Scrable and her team gave albino mice two extra genes (Cronin, C. A., Gluba, W. & Scrable, H. *Genes and Development* **15**, 1506–1517; 2001). One produces melanin pigment, allowing the mouse to grow a brown coat. But this only happens when a dietary trigger activates the regulator gene. If the sugary trigger is removed, the regulator gene switches the pigment gene off and the mouse returns to its original colour.

Scrabble hopes the technique could eventually allow endogenous genes to be activated on demand, perhaps producing better animal models of human diseases.



Mouse models: a simple change in diet switches the coat colour from white to brown.

H. SCRABLE

Brane new world

Some theorists propose that our Universe exists as a slice through multidimensional space. Could this 'brane-world' concept unify gravity with nature's other fundamental forces? Roland Pease reports.

Imagine a parallel universe in which the three familiar dimensions of space and one of time are replaced by alternative dimensions beyond our experience. Now imagine that multiple universes exist as membranes, or branes, through a multi-dimensional hyperspace. These additional dimensions could be the size of atoms, or infinitely large. We would never be able to enter them, yet they could have profound effects on the physics of our Universe.

Welcome to brane world, a bizarre territory so far only proved to exist in the thoughts of theoretical physicists. Although it might sound like a fanciful exercise in boggling the minds of lesser mortals, brane theory is in fact a serious attempt to solve the most annoying problem in modern physics: working out how to unify gravity with the other three fundamental forces of nature — electromagnetism, and the strong and weak nuclear forces.

Unequal partners

The leading explanation of how fundamental particles and forces behave is a theory called the standard model. Its main shortcoming is the difficulty of incorporating gravity on an equal footing with the other known forces. "It's remarkable that gravity, despite being the first to be discovered, is by far the most poorly understood force," says Nima Arkani-Hamed of Harvard University, one of the architects of brane-world physics.

The fundamental problem is that, com-

pared with its counterparts, gravity is weak. This can be illustrated by comparing the gravitational attraction between two electrons with their mutual electrostatic repulsion — gravity is weaker by a factor of 10^{43} . Arkani-Hamed uses a more familiar image to drive the message home: "An ordinary magnet can lift a pin off a table, even though the entire mass of the Earth is tugging down on this pin, trying to prevent it from being picked up."

Yet current theories of the Universe demand that, for a short period of time immediately after the Big Bang, all four fundamental forces were as one. Under these highly energetic conditions, before the forces split from one another as the Universe cooled, gravity must have had the same strength as its three counterparts.

The leading attempt to reconcile gravity with the other fundamental forces is string theory, in which all fundamental particles are represented as vibrating one-dimensional 'strings'. According to string theory, gravity's strength catches up with other forces if one considers its effects over small enough distances. At a length scale of 10^{-35} metres, known as the Planck length, it should be possible to probe the fundamental nature of gravity. And according to the leading version of string theory, if one were to probe space at such scales, an additional seven curled up dimensions would be revealed.

Physicists look for fundamental particles in the debris created when larger particles are



Weak link: a simple magnet is all that is needed to overcome the force of gravity.

CORE/IS

smashed together at high energies. But if conventional string theory is valid, to probe the nature of gravity would require colliders that can produce collisions with an energy of about 10^{16} tera-electron volts (TeV) — the Planck energy. This is way beyond today's most powerful machine, the newly refurbished Tevatron at Fermilab near Chicago, which operates at nearly 2 TeV.

The huge size of the Planck energy also creates problems for theorists. Experiments with particle colliders have revealed that electromagnetism and the weak nuclear force are unified at an energy of around 1 TeV — and the standard model has been carefully adjusted to fit this 'electroweak' energy scale. But it remains difficult to explain the huge difference in size between the observed electroweak scale and the 10^{16} TeV realm of string theory. Physicists call this the hierarchy problem.

Most theorists attempt to resolve the hierarchy problem by invoking supersymmetry. This theory proposes that every particle in the standard model has a counterpart differing in the fundamental property of 'spin'. The quantum effects of these particles operate at extremely high energies, near the Planck energy, but theorists have found that they almost exactly cancel out, leaving the electroweak energy scale as the residue.

In 1998, while working at Stanford University in California, Arkani-Hamed, Savas



The brains behind the branes: left to right, Savas Dimopoulos, Nima Arkani-Hamed and Gia Dvali.

Dimopoulos and Gia Dvali turned this approach to the hierarchy problem on its head. Rather than worrying about the inconvenient size of the Planck energy, they wondered what gravity would look like if it too operated at the electroweak scale, making it stronger than we realize.

Distance learning

Isaac Newton's inverse-square law states that the gravitational pull exerted by an object weakens with distance, r , according to the formula $1/r^2$. This is because, as one moves away from a gravitational source, the force's field lines spread out through three dimensions. Arkani-Hamed, Dimopoulos and Dvali started considering scenarios in which our Universe is a brane augmented by additional dimensions, into which gravity could spread. These additional dimensions are not infinite like those of space and time, but of a variety of possible sizes¹. Dimensions can be finite if they are curled up into closed loops. In essence, the Stanford researchers' proposed dimensions are similar to those of conventional string theory, only larger.

If there is one additional dimension, gravity would obey a $1/r^3$ rule over the length scales covered by this dimension. But to make gravitational effects originate at the TeV scale, this extra dimension has to be so large that its effects on gravity would alter the orbits of the planets — not really a viable option. But with just two extra dimensions, giving a $1/r^4$ law for the length scales of those dimensions, each could be less than a millimetre in size. Anomalous gravitational effects would then only show up at these distances — and that was an intriguing possibility, because physicists were only starting to measure the strength of gravity over sub-millimetre distances (see 'A matter of some gravity', overleaf).

As more dimensions are added, the length scale over which these effects would show up becomes smaller. But even with seven additional dimensions — the same number as in the leading variant of string theory — it would only shrink to about the size of an atomic nucleus. That is a far cry from the 10^{-35} m of



Crash course: the newly refurbished Tevatron will put some variants of brane theory to the test.

the Planck length, and means that it might be possible to probe the nature of gravity using the Tevatron and the next big accelerator, the Large Hadron Collider, scheduled to open in 2006 at CERN, the European Laboratory for Particle Physics near Geneva.

But why should the postulated carrier particles for gravity, known as gravitons, be able to roam through extra dimensions whereas other particles cannot? String theory has an answer. Indeed, the concept of the brane precedes the work of Arkani-Hamed, Dimopoulos and Dvali. According to current formulations of string theory, there are two types of string. The ends of 'closed' strings join up to form a loop in multidimensional space; the ends of 'open' strings do not join up, and must be stuck to a brane.

As Arkani-Hamed, Dimopoulos and Dvali showed with CERN theorist Ignatios Antoniadis², their picture works because gravitons are closed loops and so can wander through the extra dimensions. Other particles are open strings, confined to the three spatial dimensions of our brane Universe.

Warp factor

The brane-world concept has many guises, depending on how many extra dimensions are included, the size of each of those dimensions, and the different ways of wedging the models to existing theories of high-energy physics. But probably the most significant variant was devised in 1999 by Lisa Randall of Princeton University in New Jersey and the Massachusetts Institute of Technology, and Raman Sundrum, then at Boston University^{3,4}.

Randall and Sundrum were toying with brane-world theory in an attempt to iron out niggling problems with supersymmetry. They realized that branes have energy — per-

haps a significant amount — and that, according to general relativity, this energy would warp the space around it. They went on to show that this means that gravity could exist in one or more infinitely large extra dimensions without causing significant observed deviations from Newton's $1/r^2$ law. The warping caused by our brane Universe would focus gravity onto the brane and prevent it from getting far away.

In a variation on the theme, Randall and Sundrum have also shown that it is possible to envisage a pair of branes lying almost immediately side-by-side along an extra infinite dimension. One brane, with low energy, is our Universe. The other, hidden, brane is at the Planck energy. This shadow brane then acts like a sump for gravity — it is where gravitons prefer to be and is sufficiently close that they can move from one brane to the other. So, as in the Arkani-Hamed, Dvali and Dimopoulos model, gravity is far stronger than we realize but is dissipated into extra dimensions.

Other variations of brane-world theory are being used to tackle questions about the evolution of the Universe. Astronomers have evidence that there is a mysterious form of energy, known as the cosmological constant, accelerating the expansion of our Universe. Attempts to model this from first principles result in much more rapid expansion than is observed. But brane-world theorists point to a long-forgotten paper⁵ from Valery Rubakov and Misha Shaposhnikov, then at the Institute of Nuclear Research in Moscow. Crudely speaking, this paper argued that the excess energy can be drawn off into extra dimensions, in much the same way as gravity.

One entertaining version of brane-world theory offers an unusual explanation for 'dark matter', the Universe's 'missing' mass that can be detected by its gravitational influence, yet



Hidden force: Lisa Randall suggests that gravity may leak into a parallel brane universe.

Just the possibility that there could be these extra dimensions out there is so entertaining.

seems to emit no radiation. Arkani-Hamed, Dvali, Dimopoulos and Nemanja Kaloper, also at Stanford, have proposed that our brane Universe could be folded back on itself (see diagram, right), so that stars at huge distances from us along the brane could be less than a millimetre away along a higher dimension⁶. Their light would not yet have reached us because it has to travel all the way around the folded brane, but their gravity could take the short cut. If this exotic model is correct, then the Laser Interferometer Gravitational-Wave Observatory, with its detectors in Washington state and Louisiana, should detect more intense gravitational waves than are predicted by conventional theories of dark matter.

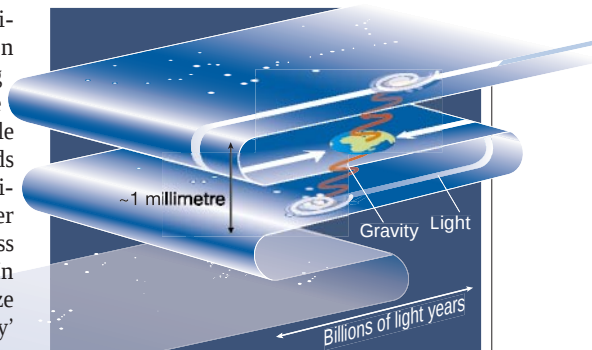
Once one allows for the existence of more than one brane, it is possible to imagine collisions during the early life of the Universe that would alter the composition of our brane. Dvali, now working at New York University, believes this could explain why our Universe

is made solely of matter and has no anti-matter, although the two should have been made in equal quantities during the Big Bang. Dvali and Gregory Gabadadze of the University of Minnesota propose that, while the Universe was young, two brane worlds collided and exchanged matter and anti-matter⁷. "What we see as an excess of matter on our brane, these other guys see as an excess of antimatter on their brane," says Dvali. In the same paper, Dvali and Gabadadze describe an alternative scenario of 'baby' branes being boiled off our own brane during the Big Bang, carrying away antimatter.

Fertile environment

The sheer fecundity of brane-world theory is part of its appeal. But its critics complain that the multitude of possibilities it throws up makes it impossible to pin the brane-world theorists down. "The whole approach is not falsifiable," says Gordon Kane, a theoretical physicist at the University of Michigan in Ann Arbor.

Nevertheless, at least some of the predicted brane-world scenarios are now being tested. The range of possibilities, in fact, has already been constrained by observations of Supernova 1987A, which shows that little of its energy could have been released into higher dimensions⁸. That effectively limits the size of the extra dimensions to the sub-millimetre scale. And precise investigations of gravity



Cosmic origami: a folded brane Universe would allow gravity to slip through an extra dimension.

over these distances are now probing for the extra dimensions predicted by Arkani-Hamed and his colleagues (see 'A matter of some gravity', below left).

Meanwhile, high-energy physicists are looking for anomalous results in their experiments that might suggest that gravitons are disappearing into and emerging from the extra dimensions. The most stringent test so far was published in February by the D0 group, working at Fermilab⁹. Researchers led by Greg Landsberg of Brown University in Providence, Rhode Island, re-analysed 60 million particle collisions recorded at the Tevatron between 1992 and 1996, looking for evidence that more pairs of photons, or paired electrons and positrons, were being produced than predicted by the standard model. Additional pairs might be generated by gravitons emerging from an extra dimension, Landsberg explains. His team's negative results suggest that if there are two extra dimensions, they can be no larger than 0.3 millimetres; if there were as many as seven, none can be larger than 2 femtometres (2×10^{-15} m).

The revamped Tevatron, and future colliders, should be able to constrain the brane-world possibilities still further, and may also provide evidence for supersymmetry. But even if these experiments demolish the brane-world's attempts to rival supersymmetry, its enthusiasts seem unlikely to abandon their multidimensional thought experiments. "Just the possibility that there could be these extra dimensions out there is fundamentally so entertaining," says Randall. "It's fun to think about."

Roland Pease is a science producer for BBC Radio.

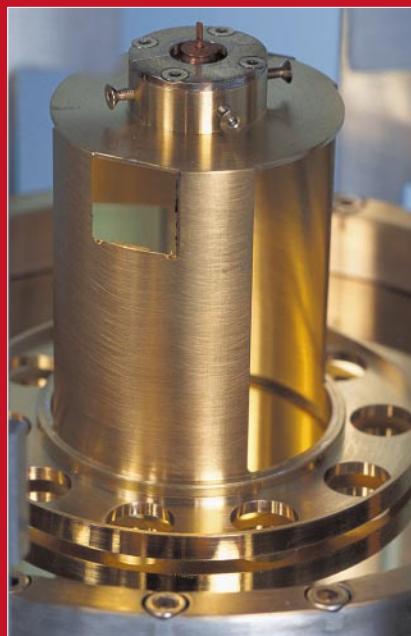
A matter of some gravity

Eric Adelberger and Blayne Heckel of the University of Washington in Seattle are no strangers to difficult gravity experiments. In the 1980s, they led one of a number of groups that investigated the existence of a postulated fifth force, which would show up as a gravitational anomaly over distances of up to 100 metres. Their findings¹⁰ helped to kill the idea.

So when they heard about the possible existence of extra dimensions, and that these might reveal themselves as deviations from classical newtonian gravitation over distances of less than a millimetre, it was natural that they would pursue them.

The apparatus they use is a marvel of precision engineering. It consists of a pair of metal discs, one above the other (see right). The upper, aluminium disc is suspended on a fibre of tungsten; the lower, copper disc is spun by a precise motor. A series of holes cut around the perimeter of each disc means that when the lower disc is rotated slowly it exerts little effect on the upper disc with purely newtonian gravity, but any deviations resulting from extra dimensions will cause a just-detectable twist on the upper disc.

The angular movements needed to signal the presence of additional dimensions are incredibly small — just a millionth of a degree. In February, Adelberger and Heckel reported that they could find no evidence for extra dimensions over length scales down to 0.2 millimetres (ref. 11). But the quest goes on. The researchers are now designing an improved instrument to probe the existence of extra dimensions below 0.1 mm. Other physicists, such as John Price of the University of Colorado and Aharon Kapitulnik of Stanford University in California, are attempting to measure the gravitational influence on small test masses of tiny oscillating levers.



1. Arkani-Hamed, N., Dimopoulos, S. & Dvali, G. *Phys. Lett. B* **429**, 263–272 (1998).
2. Antoniadis, I., Arkani-Hamed, N., Dimopoulos, S. & Dvali, G. *Phys. Lett. B* **436**, 257–263 (1998).
3. Randall, L. & Sundrum, R. *Phys. Rev. Lett.* **83**, 3370–3373 (1999).
4. Randall, L. & Sundrum, R. *Phys. Rev. Lett.* **83**, 4690–4693 (1999).
5. Rubakov, V. A. & Shaposhnikov, M. E. *Phys. Lett. B* **125**, 139–143 (1983).
6. Arkani-Hamed, N., Dimopoulos, S., Kaloper, N. & Dvali, G. *J. High Energy Phys.* [online] <<http://jhep.sissa.it/archive/papers/jhep122000010/jhep122000010.pdf0012>> (2000).
7. Dvali, G. & Gabadadze, G. *Phys. Lett. B* **460**, 47–57 (1999).
8. Hannestad, S. & Raffelt, G. G. *Phys. Rev. Lett.* (in the press).
9. Abbott, B. et al. *Phys. Rev. Lett.* **86**, 1156–1161 (2001).
10. Stubbs, C. W. et al. *Phys. Rev. Lett.* **58**, 1070–1073 (1987).
11. Hoyle, C. D. et al. *Phys. Rev. Lett.* **86**, 1418–1421 (2001).

Vital signs

The science of biomonitoring, which uses living organisms as 'sensors' to track environmental pollution, seems to be coming of age. John Whitfield considers its potential.

In the early 1980s, an illegal battery-disposal operation in Hong Kong's Junk Bay was releasing large amounts of polychlorinated biphenyls, lead and zinc into the water. But the crime did not go unwitnessed. The barnacles and mussels living in the bay concentrated the pollutants in their tissues. The evidence they gave up to local researchers and their colleagues at the Natural History Museum in London helped the authorities shut the law-breakers down.

The idea that studies of living organisms can provide information about environmental hazards is not new: before the advent of modern safety equipment, miners kept an eye on the health of caged canaries to warn them of dangerous gas build-ups. But as researchers have concentrated on their own favoured techniques, rigorous standardized methods for biological monitoring have been slow to emerge — success stories like the Hong Kong example are still rare.

Enthusiasts for biomonitoring argue that their field is now coming of age, however. They point to the recent development of protocols that can do much more than simply provide general markers of ecosystem health. In many cases, researchers are now combining ecological studies with analytical chemistry to produce information on the effects of pollution on living organisms, the identity of the chemicals involved, and even where they came from. "In the past few years there's been a considerable drive on the part of the main regulatory bodies to integrate biological and chemical monitoring," says Peter Matthiessen, an ecotoxicologist and director of the UK Natural Environment Research Council's Centre for Ecology and Hydrology in Windermere, Cumbria.

Biomonitoring has long been the poor relation to the straightforward chemical analysis of water, air and soil. Chemical sensors can provide highly accurate readings of environmental pollution. But in some



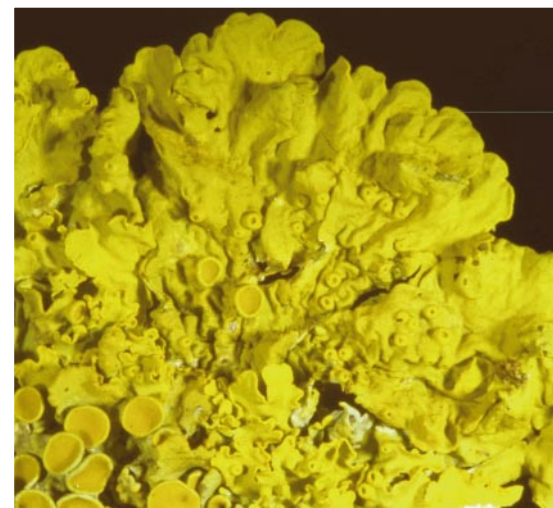
Telling tales: *Mytilus edulis* mussels (above) and lichens such as *Xanthoria* (right) accumulate chemical pollutants in their tissues.

regards, say the proponents of biomonitoring, this precision is spurious. Instruments can quantify the amount of a pollutant present in the environment. But if a pollutant is not taken up by organisms, it may cause little damage to an ecosystem — and the extent to which it is taken up may depend on a range of factors, including climate and acidity. Also, chemical sampling of the environment can only provide a snapshot of what may be a highly dynamic situation, whereas some organisms preserve a continuous record of the environment throughout their lives.

Community values

In the early years of the last century, two German biologists, Richard Kolkwitz and Maximilian Marsson, realized that some freshwater invertebrates were more sensitive to pollution than others — which means that the community of species found at a particular site says much about its cleanliness. In Britain, this technique is used to monitor 7,000 river sites across the country.

Although assessments based on community ecology are good at exposing severe pollution events, they are less useful at providing subtle warning signs that an ecosystem is coming under pressure. "They can only tell you that you've had a major impact after the event," says Matthiessen. But combining ecological observations with chemical measurement of pollutants accumulated by animals



and plants can provide a much more sensitive and predictive analysis.

Marine scientists have led the way. In Europe, their work was stimulated by the 1992 Convention for the Protection of the Marine Environment of the North-East Atlantic, known as the OSPAR Convention, to which most European nations are signatories. The OSPAR Convention covers the northeast Atlantic, North Sea and parts of the Arctic Ocean and Mediterranean. Its signatories have pledged to "take all possible steps to prevent and eliminate pollution".

A requirement to monitor the effects of pollutants — including heavy metals, industrial and agricultural chemicals, radioactive waste — and the activities of the oil and gas industries on marine organisms is written into the OSPAR Convention. But as measurements began to accumulate over the 1990s, it

► became apparent that the diversity of methods used was preventing useful comparisons.

In 1998, the year in which the OSPAR Convention came into force, the European Union set up a project called Biological Effects Quality Assurance in Marine Monitoring, or BEQUALM, to standardize marine biological monitoring. By the time this project wraps up in October, a network of laboratories in Britain, Norway, Sweden and Germany should have hammered out about half-a-dozen standardized measures. These include the activity levels in fish of enzymes that process trace metals and organic pollutants; pathological analysis of fish livers; and community analysis of planktonic plants and invertebrates living in the seabed. Given the range of techniques involved, the progress towards consensus is no mean achievement, says Matthiessen.

Researchers working on other aspects of biomonitoring are similarly standardizing. "There was a period of total anarchy, when every scientist had his or her own methods," says Pier Luigi Nimis, a botanist at the University of Trieste in Italy who uses lichens to monitor air pollution. He now believes the best methods are emerging "by a process of natural selection".

Italian lichenologists have adopted a dual approach. They have devised an index of lichen biodiversity, and the sampling methods to calculate it, as an indicator of the atmospheric levels of sulphur dioxide and oxides of nitrogen. Coupled with this, the accumulations of 17 trace metals are measured in a single species in each area.

Governments are starting to take notice. ANPA, the Italian environment agency, has launched a lichen-mapping project. And the influential Association of German Engineers intends to submit a slightly modified version of the Italian protocols to the European Committee for Standardization in Brussels for adoption at a pan-European level.

Enthusiasts say biomonitoring is much cheaper than conventional chemical monitoring. Automated chemical sensors are expensive to buy and maintain, says Nimis, and biological approaches can help reveal where best to use them: for example, by revealing pollution hotspots that merit continuous chemical monitoring. "Moving 500 metres can make a big difference," he says.

Steve Hopkin, a zoologist at the University of Reading who is trying to interest the

After a period of total anarchy, the best methods are emerging by a process of natural selection.



Under surveillance: barnacles and mussels have been used to monitor pollution in Hong Kong's waters.

Environment Agency of England and Wales in biomonitoring protocols involving earthworms and woodlice, adds that instruments are vulnerable to theft and vandalism. "If you try to set up an air sampler in an inner city, it's not going to be there for very long, unless you put an electrified fence around it," he says.

Star quality

As biomonitoring slowly gains credence with regulatory agencies, certain 'indicator' species have emerged as stars. In the sea, the undisputed champion is the blue mussel, *Mytilus edulis*. As mussels filter food from the water they live in, they also retain contaminants, which reach high concentrations in their tissues. Their sedentary lives prevent confusion about where they might have picked up a chemical. They are found in vast numbers all around the northern hemisphere, and they are an important food source for many animals.

Several countries have mussel-watch programmes designed to reveal large-scale, long-term trends. The US Mussel Watch programme, run by the National Oceanic and Atmospheric Administration, was established in 1986. It monitors mussels from 263 sites around the US coastline, focusing on trace metals and organic compounds such as polychlorinated biphenyls and dioxins. Since it began, levels of most synthetic chemicals and of cadmium have fallen, whereas other trace metals have held steady.

In warmer climes, where *M. edulis* does not live, crustacea could become the sentinel organisms of choice. Barnacles in particular, says Philip Rainbow of the Natural History Museum, are "phenomenal accumulators of trace metals". Rainbow advocates the use of 'cosmopolitan' crustaceans with wide distributions, such as the barnacle *Balanus amphitrite*, which has spread around the world by clinging to the hulls of ships.

Rainbow is one of the leaders of the project to monitor barnacles and mussels in the waters of Hong Kong. In addition to fingering specific polluters, the project has recorded the shifting pattern of industry and pollution in the former British colony. Over the 1980s, the focus of pollution moved from Victoria Harbour in the south to Tolo Harbour in the north. In the 1990s, as Tolo Harbour was cleaned up, the distribution of pollutants began to reflect the growing industrialization of the neighbouring part of China.

Valuable though 'indicator' species such as mussels and barnacles are, studies of just one or two organisms cannot reveal everything about an ecosystem. To gain a complete picture of the marine environment, information from a filter feeder such as a mussel should be augmented with analysis from a seaweed, which samples chemicals in solution, and a sediment-dweller such as a worm.

Although most biomonitoring experts are optimistic that the use of such wide-ranging protocols will increase, they note that many regulators are still more comfortable with chemical analyses — which tend to be easier to enforce, and for courts to interpret in the event of breaches of pollution controls. "These things don't change overnight," says Matthiessen. But now that biomonitoring has its foot in the regulators' door, he is confident that its day will come. ■

John Whitfield works in Nature's science writing team.

OSPAR Convention

♦ <http://www.ospar.org/eng/html/convention/welcome.html>

BEQUALM

♦ <http://www.cefas.co.uk/bequalm>

Italian lichen protocols

♦ http://www.sinanet.anpa.it/aree/atmosfera/qaria/biomonitoraggio/Nimis_Bioindi.asp

US Mussel Watch

♦ http://state-of-coast.noaa.gov/bulletins/html/ccom_05/ccom.html

'Espionage' charge may be based on a misunderstanding of rules

Researchers working abroad need to know local laws

Sir— Differences between the perceptions of Japanese and US academics as to who owns and controls the fruits of academic research may lie behind the events leading to the indictment of two Japanese researchers for violating the US Economic Espionage Act and for interstate transportation of stolen property (*Nature* **411**, 225–226; 2001). The indictments alleged that, in leaving the Cleveland Clinic to take a position at the Institute of Physical and Chemical Research (RIKEN) near Tokyo, Takashi Okamoto transferred cells and reagents to his friend Hiroaki Serizawa's laboratory at Kansas University Medical Center, then returned to Kansas to take the cells and reagents to Japan.

Although reports in the Japanese media suggest that the indictment against Okamoto and Serizawa reflects increased Japanese–US competition in biomedical research, the underlying reasons may be both more mundane and more complex. Just over two years ago, another Japanese researcher was charged with stealing research data from the Mayo Clinic. Was nationalistic rivalry the driving force behind both incidents? Are Japanese researchers prone to thievery at the behest of their home institutions? Are US academic research centres and the FBI inclined to intimidate the Japanese researchers who devote some of their most productive years to working (at relatively low pay) on research in these centres? The answer to these questions is probably “no”.

In Japan, university researchers are considered to own most of the fruits of their research, including copyrightable work, most patentable inventions (with the rare exception of inventions designated as ‘national’), data, and non-patented biomaterials such as cell lines and diagnostic reagents. Although there is no government regulation specifying who should own or control biomaterials produced in Japanese university laboratories, the accepted practice is for the inventing professor to control them. When inventors move to other institutions, they usually take such materials with them, unless the project is governed by special regulations or there is a contract prohibiting this practice.

In contrast, most US academic research institutions require their researchers to assign most of the fruits of their research to their institutions as a condition of

employment, including patentable inventions, any material inventions (including non-patented biomaterials), copyrightable computer programs, and tangible data. Although this may seem avaricious to Japanese researchers, there are three main reasons for these policies.

First, the 1980 US Bayh–Dole Act gave universities the right to claim ownership over all potentially patentable inventions made with government funding. This act was designed to give universities incentives to ensure the development and commercialization of the fruits of their research. It is easier for universities to exercise their authority under Bayh–Dole and to coordinate the development of various discoveries, if they control all tangible products of research.

Second, institutional ownership of research materials and data provides for promising research to continue even if the inventor leaves. US academic researchers frequently move, usually agreeing to transfer materials to their new laboratories so that they can continue their research. Ownership is a safeguard if a move is likely to mean the results would otherwise be lost to the academic community.

Third, the university–researcher relationship in the United States is much closer to a private-sector employer–employee relationship than that between Japanese national universities and their staff. Even in US state universities, researchers are employees of their universities, not of the state, and they owe their universities much the same loyalty and rights over the products of their work as do employees of private companies.

The fact that US academic institutions own research material does not mean that they are free to restrict its use by others. Except in the case of inventions that have commercial potential and need substantial additional investment to make them marketable, the National Institutes of Health (NIH) encourages the free transfer of non-patented biomaterials for research purposes. Under NIH guidelines, such transfers should occur under material transfer agreements (MTAs) that place minimal restrictions on the recipients' use of the materials, provided the use is for non-commercial purposes and not in humans. Hence, Okamoto should in principle have been able to receive copies of the cell lines and reagents he made while

at the Cleveland Clinic via an MTA between the Cleveland Clinic and RIKEN.

However, MTAs are still uncommon in Japanese academic research. Therefore, it is likely that Okamoto did not realize the implications of the Cleveland Clinic's ownership of his materials, and did not know that he probably could have taken them legally via an MTA. In any event, increased awareness of these issues will help avoid incidents such as this one.

Robert Kneller

Research Center for Advanced Science and Technology, University of Tokyo, 4-6-1 Komaba, Meguro-ku, Japan

You can help rare plants survive in the cities

Sir— Almost a quarter of the endangered plant species in the United States (listed at <http://www.abi.org>) are in densely populated urban areas. We investigated the status of plants at risk in one of these, the metropolitan San Francisco Bay area, to assess the degree to which endangered urban plants are protected.

The California Natural Diversity Database (CNDDB, at <http://www.dfg.ca.gov/whdab/html/cnddb.html>) identifies 68% of the 214 listed species at risk in this area as having at least one population on publicly owned land. However, there is no centrally organized effort to ensure adequate protection of plant species in the area.

We surveyed 43 statutory authorities that owned land, and found that fewer than half monitor any plant species; only 10% could assess trends in population size for any species. Lack of funds is the main reason for not managing rare plants.

The lack of attention to stewardship of endangered plants in the urban environment needs to be rectified. Although small populations are prone to high extinction risk, there are few case studies documenting extinction in small, naturally isolated populations (Schwartz, M. W. *Annu. Rev. Ecol. Syst.* **30**, 83; 1999). In contrast, there are numerous examples of small-plant population persistence. For example, a population of *Streptanthus niger* or Tiburon jewelflower, a rare annual restricted to California's Tiburon peninsula, has persisted under benign neglect within an urban landscape since at least 1945 (see CNDDB website).

The high number of rare plants in urban areas that need monitoring and management attention presents an ideal opportunity for volunteer involvement. All 6.9 million residents of the San Francisco Bay region live within eight miles of a plant species at risk, and an overwhelming

majority of the landowners we surveyed welcomed the idea of volunteer assistance in managing rare plants.

Effective conservation requires public support. Volunteers could provide assistance by removing weeds or by monitoring population size. In order to be effective, adopt-a-plant volunteer programmes need to coordinate mechanisms for systematically collecting, archiving and analysing data.

Alerting the public to how they can contribute to the preservation of biodiversity in urban environments can keep people interested in the natural world in which they live, and just might save some species from extinction.

M. W. Schwartz, N. Jurjavcic, J. O'Brien

*Department of Environmental Science and Policy,
University of California, Davis,
California 95616, USA*

Undergraduates must learn to communicate ...

Sir — We strongly support the argument of your Opinion article "Learning to speak and write" (*Nature* **411**, 1; 2001). Among the most important areas to be mastered by science students is the ability to present findings orally to one's peers. Learning to do so can be essential in order to function as an independent scientist.

After a pilot study indicating that our graduate students achieved little or no improvement in presentation ability after taking our graduate-student seminar, we redesigned our course curriculum. Six students reviewed a recent scientific paper in a field of their interest, in three 30-minute oral presentations. Their brief was to provide background information to clarify the rationale; succinctly but unambiguously state the hypothesis to be tested; discuss the experiments in the paper; briefly detail the results obtained; outline the paper's conclusions; and critically assess the quality of the work.

Our goal was for this exercise not only to help students read and evaluate the scientific literature critically, but also to help them become more adept at orally presenting information to their peers.

Our idea was that through each of the three presentations, students should become more polished in both oral presentation and critical analysis. Before the first presentation they each had three sessions with a staff mentor, including a 'dress rehearsal'. They had two sessions before the second presentation and just one brief meeting before the third. Each student and staff reviewer was asked to rate the presenter according to how effectively information was presented in the six

categories detailed above.

The overall mean presentation score awarded by both students and staff evaluators improved significantly from presentations one to two ($P < 0.001$ and $P < 0.05$, respectively). The student evaluators gave roughly the same mean scores for the third presentations (prepared with very little help) as for the second. Staff, however, were less impressed by the third than the second presentations, and gave scores significantly lower than those of student evaluators ($P < 0.001$).

Even so, this small study indicates that students' oral presentation skills improve on a curriculum that requires a number of seminars prepared with decreasing help. The fact that by the third round students could independently deliver a presentation as well-received by their peers as their first efforts (which were achieved with substantial help) indicates strong progress. Whatever the reasons for the variation in grading of the third presentations, staff and students agreed that these were clearly better than the first talks. From a faculty perspective, we believe the improvement was well worth the effort.

**David A. Watson, Brian R. Davis,
David W. Niesel**

*Department of Microbiology & Immunology,
University of Texas Medical Branch, Galveston, and
InDyne, Inc., 18108 Point Lookout Drive, Suite B,
Nassau Bay, Texas 77058, USA*

... following the lead of pioneering teachers

Sir — I was surprised by your Opinion article¹ calling for science-communication skills to be taught as part of undergraduate degrees. My colleagues and I had assumed such courses were common. The development of scientific communication skills is an integral part of our degrees in the department of biological sciences here. The compulsory formal training element represents a sixth of the undergraduate degree, over three years. It was highlighted for praise during our recent assessment by the UK quality-assurance agency.

Our course is focused on progressive skills, methods and concepts modules in which students receive training in general communication (essay and paper writing, oral presentation, abstracting) and skills directly associated with their chosen degree subject. The second-year methods modules, for example, include rigorous training in experimental design and data analysis. The culmination of training in the third (final) year includes scientific debate, a major literature review, the construction and presentation of scientific posters, and individual and group

presentations using latest technology.

Marine biology and ecology students write up their honours research projects as a manuscript for a relevant journal, following the journal's instructions to authors. All students also present their research project at a three-day conference attended by staff and students, including second-year students in the process of choosing their dissertation topic.

The work of a significant number of our marine biology and ecology students has been published in the primary literature (see, for example, refs 2 and 3) and major conferences, or has been used as the basis of publications, including a Letter to *Nature*⁴. As such formal training appears, from your Opinion article, to be rare, perhaps our department's courses could provide a model for degrees elsewhere.

Martin Attrill

*Department of Biological Sciences, University of
Plymouth, Drake Circus, Plymouth PL4 8AA, UK
m.attrill@plymouth.ac.uk*

1. *Nature* **411**, 1 (2001).
2. Wildgust, M. A. & Jones, M. B. *Aquat. Toxicol.* **41**, 187–192 (1998).
3. Bowden, D. A., Rowden, A. A. & Attrill, M. J. *J. Exp. Mar. Biol. Ecol.* **259**, 133–154 (2001).
4. Sims, D. W. & Quayle, V. A. *Nature* **393**, 460–464 (1998).

Brazil's WTO case has no bearing on AIDS

Sir — Your Opinion article "Next steps against AIDS" (*Nature* **410**, 1009; 2001) misstated the facts about how Brazil is coping with the epidemic. It also mischaracterized the views of the Pharmaceutical Research and Manufacturers of America (PhRMA).

Brazil has never utilized compulsory licensing for any patented pharmaceutical product. Indeed, 80% of the medicines distributed to the 90,000 AIDS patients under treatment are purchased directly from research-based pharmaceutical companies. Many are members of PhRMA. The remaining 20% of drugs used in the Brazilian programme are made locally and are not covered by patents.

PhRMA is on record as a supporter of the Brazilian model for prevention and treatment of HIV/AIDS. The case before the World Trade Organization (WTO) has no bearing on Brazil's AIDS programme. It focuses on the 'local working' requirement of Brazil's patent law, which discriminates against imported products and violates the WTO agreement. The dispute centres on a protectionist barrier to trade in all intellectual-property-related products.

Alan F. Holmer

*Pharmaceutical Research and Manufacturers of
America, 1100 Fifteenth Street NW, Washington,
DC 20005, USA*

Who is out there?

Search for Life/Astrobiology

by Monica Grady

Natural History Museum/Smithsonian

Institution: 2001. 96 pp. £9.95/\$14.95 (pbk)

Don Brownlee

Living in the thin but luxuriant biosphere of Earth, it stretches the mind to think that life would be difficult to find elsewhere. Only a century ago it was possible to imagine that creatures similar to humans might live on our neighbouring planets, but a century of investigation has yielded no totally convincing evidence for extraterrestrial life. It surely exists, but it has not been easy to find. The search is complicated by the probability that

environments as favourable as Earth's are not common in the Universe, and it is likely that life elsewhere will not be nearly as obvious as it is in our fecund biosphere. Instead of looking for people, Klingons or Wookies, the search will probably have to concentrate on detecting bacterium-like organisms similar to those that have dominated more than 80% of the history of life on Earth.

The fascinating and richly interdisciplinary story of the search for extraterrestrial life is admirably summarized by Monica Grady. The search reached a watershed, almost its Waterloo, in 1976, when the two Viking missions travelled to Mars to search for life there. Although the two landers and orbiters were huge successes, they found no evidence of life, and the data they sent back were widely interpreted as evidence that life, or even organic compounds, could not exist on the martian surface. Mars, the most Earth-like planet in the Solar System, was in some ways more hostile to life than the surface of the Moon, dashing decades of hopes of a breakthrough.

Those decades before Viking had seen the birth of the science of exobiology, the instigation of searches for radio signals from other intelligent life, the formulation of the Drake equation (a means of estimating the number of civilizations in the Galaxy), the discovery of interstellar organic molecules, the abiological synthesis of amino acids and the discovery of extraterrestrial amino acids in the Murchison meteorite. Enthusiasts hoped that exobiology might soon give us exciting news of the nature and abundance of life on other planets. Nay-sayers complained that exobiology was not a science and never would be, because it had no data and never would have any. Detractors simply restated Fermi's paradox about the existence of extraterrestrial life: "So? Where is everybody?"

A quarter of a century after Viking, exobiology has found a new incarnation as astrobiology — a field that has expanded to cover almost all aspects of the origin and evolution of life on Earth, and the search for suitable habitats and life elsewhere. Although alien life has remained, and may always remain, elusive, a wealth of recent discoveries has given vigour and optimism to the field. The discoveries of life in deep-ocean thermal vents, in rocks many kilometres below Earth's surface, in fluid inclusions in ice and in dry Antarctic valleys has extended our idea of the environmental conditions in which it can occur on Earth. Beyond Earth, an apparent ocean has been discovered on Europa, one of Jupiter's moons, and there is evidence for ancient water on Mars, along with controversial hints of microfossils in martian meteorites. Other developments include the discovery of planets beyond our Solar System as well as new understanding of the nature and evolution of terrestrial life and the causes of global extinctions.



Life at the extremes

The discovery that life can exist on Earth in conditions previously thought impossible has raised hopes that it could exist elsewhere in the

Universe. Above, the Minerva Terrace hot springs in Yellowstone National Park, from *Life Beyond Earth* by Timothy Ferris (Simon & Schuster, \$40).

Astrobiology is a rich intersection of many fields of science that include microbiologists, geologists, atmospheric scientists, astronomers, oceanographers, glaciologists and chemists, who frequently meet to discuss the nature of life in the cosmos. Whether or not we ever find alien life, astrobiology has landed and established a permanent beachhead in academia. *Search for Life* is an excellent introductory overview of its current status, covering the origin of known life and its nature and the search for it elsewhere. The book is not intended to cover the field in depth, but is a well-illustrated, whirlwind tour of most of the significant areas. Nevertheless, I imagine that even the most experienced astrobiologist will learn something from it. The broad expertise involved in astrobiology is always a challenge, and even 'experts' will benefit from such an informative overview. ■

Don Brownlee is in the Department of Astronomy, University of Washington, Seattle, Washington 98195, USA.

More astrobiology

Life Everywhere: The Maverick Science of Astrobiology

by David Darling
Basic Books, \$26

The view from Japan

Primate Origins of Human Cognition and Behavior

edited by Tetsuro Matsuzawa
Springer, 2001. 587 pp. \$119

Marc Bekoff and Jane Goodall

The past decade has seen enormous growth in the study of animal cognition. Anecdotal reports and detailed observational experiments abound for many species and for diverse cognitive skills. But much of the research on primate cognition has been done in Japan, and valuable information, not published in English, has been poorly circulated. Now, for the first time, detailed accounts of Japanese research on primate cognition have been compiled in this excellent volume. Tetsuro Matsuzawa summarizes and synthesizes previously published data as well as presenting new information.

The book is refreshing for the way in which it shows that detailed, single-subject case studies — for example, early vocal development in a chimpanzee infant — and anecdotal accounts are very important for advances in cognition and for informing subsequent experiments. In addition, it proposes new techniques for studying old problems; such methods hold much promise for future comparative research.



Another thinker? Whether primates other than humans have a notion of 'self' is an intriguing question.

'Animal cognition' refers to the many ways in which individuals sense and interact with the world. In seven sections and 28 chapters, *Primate Origins* considers some 90 species, including humans. Topics are discussed from evolutionary, comparative and ecological perspectives. They include perception and cognition, speech and language, learning and memory, recognition of self and others, imitation, tool use, social behaviour and social organization, and culture.

We are told that the human ability to handle many social relationships and different 'levels' simultaneously is related to our massive cortex. 'Levels' range from situations in which objects are dealt with singly to those in which there are several different possible relationships between them, for example a nut to be cracked, an anvil and a stone tool. The ability of early humans to perceive the world in terms of the relationships between three things (self, other members of the species and objects) underpins the advanced technology and spoken language that characterizes *Homo sapiens*.

To identify how cognition varies between species, we must understand the historical and adaptational factors that have influenced the evolution of information processing. For instance, rhesus monkeys, chimpanzees and chickens (but not pigeons) perceive a figure as a single entity even if it is hidden by an object and later emerges. It is not the size of an animal's brain, but rather the nature of its food, that appears to be responsible for the evolution of this ability. Some species need to be able to track moving food, whereas others do not.

Chimpanzees process information about light and shade differently from humans, and this seems to be related to their tropical-forest habitats. Ecological variables can also dictate species differences in vocal communication and spatial memory. Furthermore, animals of many species need to be able to estimate the number of individuals in a group or the amount of food available in a certain locale. Studies show that many primates are very adept at such 'counting', estimating and object-categorizing.

A major question in primate cognition

centres on the notion of 'self' — are individuals 'self-aware', do they possess 'self-consciousness', and what does it mean to make such claims? Much research follows Gordon Gallup's 'mark' test, in which an animal is anaesthetized, a mark is placed on its forehead that can only be seen in a mirror, and the response to the mark is assessed when the animal wakes and is shown its reflection in a mirror. A hand movement towards the mark is taken to suggest self-recognition.

But not all primate species are suited to this experiment. Thus, although squirrel monkeys don't show evidence of self-recognition in a mirror, this does not mean they are devoid of all self-awareness. Furthermore, it is not clear that self-directed movements indicate that the animal possesses a sense of self or a sense of its own body. Given that many species have never been studied and that problems with certain analyses still remain, it is premature to draw up taxonomic borders for species differences in 'self-awareness'. Nonetheless, there seems to be a qualitative gap in self-recognition between hominoids and other primates.

There is still much to be learnt about primate cognition. Without an understanding of perceptual and cognitive skills, we cannot appreciate the elaborate patterns of social behaviour to be found in diverse species, or individual and within-species variation. This understanding is also essential for studying the evolution of sociality, social transmission and culture, and the mental processes that underlie the ability to read other individuals' mental and emotional states.

Japanese scientists have played a major role in the development of primatology as a scientific discipline. For 50 years they have studied the social behaviour of wild monkeys on Koshima, while chimpanzees in Tanzania and Guinea have been observed since the mid-1960s. Culture in animals was first described by Japanese scientists studying sweet-potato washing by Koshima monkeys. There are now many examples of culture in primates, and the finding of some 40 different behaviour patterns in chimpanzees — in, for example, tool use, grooming and patterns of courtship — shows that cultural variation exists in these animals.

The many significant contributions to animal cognition made by Japanese researchers can now be appreciated by scientists outside Japan. Tetsuro Matsuzawa has done a masterly job of editing this volume, even down to the fine details of organization and style. The book will have a major impact on future studies of animal cognition in all species. ■

Marc Bekoff is in the Department of Environmental, Population, and Organismic Biology, University of Colorado, Boulder, Colorado 80309-0334, USA. Jane Goodall is at the Jane Goodall Institute, Silver Springs, Maryland 20911-4890, USA.

Science in culture

Rhythms and reflections

Life represented in architecture by the artist Andreas Horlitz.

Alison Abbott

It is a hollow steel column, lit from inside by neon tubes. The neon light shines through cut-out horizontal bars, which climb like the asymmetric rungs of a ladder up its dramatic 25-metre length. It stands in the atrium of the 1997 Gerling building in Düsseldorf, rising through all of its six storeys. At night its beam can be seen through the building's glass walls.

As part of the German movement known as *Kunst am Bau* — art within architecture — which is currently enjoying a renaissance, the column was designed by the Munich-based artist Andreas Horlitz.

Horlitz is perhaps the first to apply scientific motifs to *Kunst am Bau*. In this case, his motif is circadian rhythms. The column's horizontal bars represent seven years in the life of Alexander Borbély, a Swiss sleep researcher who has been recording his daily activity patterns since 1982 using a monitor strapped permanently to his wrist.

Each round of the column represents 48 hours. The open bars are times of low activity — sleep. The vertical shifting of the bars represents Borbély's trips to the United States and to Japan. The column could thus be seen as a monument to Borbély's life.

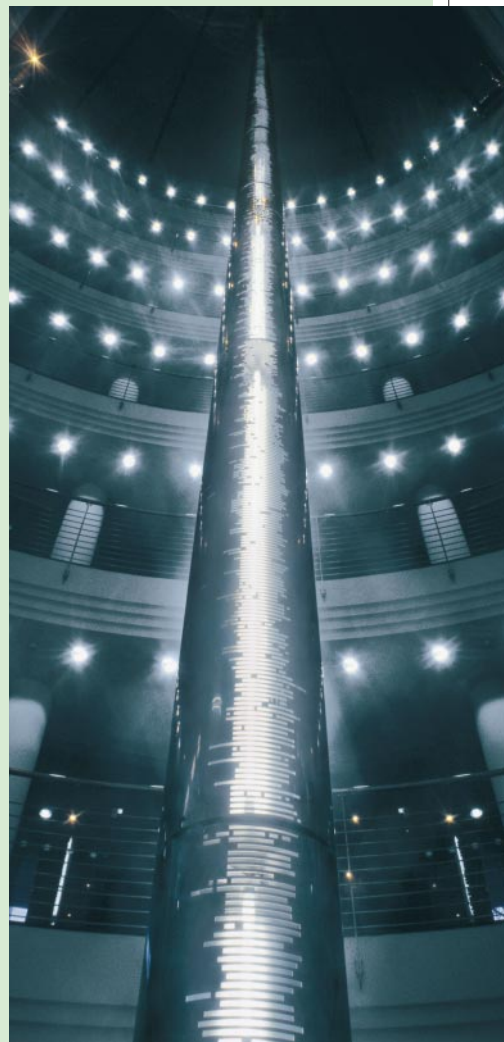
But Horlitz does not see his work in personal terms. Life's fundamental rhythms, like life's fundamental building-blocks, can be generalized, he says. He turned to circadian rhythms a few years ago, when a chance meeting with a circadian scientist in Munich's fashionable Schumann's American Bar introduced him to one of science's fashionable themes. Before that, he had spent many years working with the patterns of the biotechnologist's classical sequencing gels. In these gels, DNA fragments are separated along four tracks, one for each of DNA's constituent nucleotides, to determine gene sequence, and each fragment is seen as a spot.

DNA is a well-rehearsed theme in art. But Horlitz's approach, in which the spots are typically represented as mirroring overlaid on sand-blasted glass, has originality in that it plays on ideas of self-image. Horlitz is intrigued by the fact that individual variations in the human genome are so small, and sees his mirrored representations of a single sample of DNA as being universal. In his work, the mirroring reflects back the viewer's face while at the same time depicting a different type of self-portrait.

Horlitz does not accept commissions from biotechnology companies, which would use his ideas decoratively to represent their products. He wants his work — the representation of life at its most fundamental levels — to be part of the physical structures within which ordinary life is

played out. Panels of his part-mirrored, translucent 'DNA' glass have been incorporated into the internal walls of a commercial building — Uniplan International — in Kerpen. Gerling is an insurance company. ■
Alison Abbott is senior European correspondent of Nature.

ANDREAS HORLITZ



Archaeology of type

Printing technology co-evolved with the written representation of language.

Blaise Agüera y Arcas and
Adrienne Fairhall

Before the emergence of printing in the mid-fifteenth century, the representation of the written word in Europe was extremely complex. Especially in Germany, Gothic aesthetics and the cost of vellum (made from sheepskin or calfskin) had compressed writing on the page into the densely woven *textura* style. Abbreviated word forms abounded, and the alphabet itself had been expanded with hundreds of space-saving symbols. A few examples, such as æ, œ, & and the Spanish ñ (short for 'nn'), survive.

The arrival of paper in Europe removed the economic constraint of vellum. First announced in 105 AD in China, papermaking slowly made its way along the Silk Route, reaching Germany in 1390, when a papermill opened in Nuremberg — less than 200 kilometres from Mainz, where Johann Gutenberg would set up the first European press in about 1450. Printing was soon to play a central role in kindling Renaissance humanism, and in catalysing the Protestant Reformation by bringing the Bible and the tracts of its interpreters to a new lay readership. It is difficult to imagine the Enlightenment and the scientific revolutions of the next centuries without this technology, which allowed ideas to be disseminated quickly, broadly and reliably.

The idea of assembling a composite printing surface from small, reusable (or *moveable*) pieces of type had been developed centuries earlier in the Far East, but unlike papermaking, there is no evidence of a slow diffusion of this technology to Europe. Printing appears to have evolved independently several times; all modern printing, however, derives from Gutenberg. In distinguishing his invention from earlier Chinese and Korean printing, most scholars cite the introduction of the font: the set of unique steel master letters, called punches, used to strike matrices, from which lead letters, or types, were cast in large numbers using an adjustable mould. This underlying multiplicative process, not of printed pages but of metal types, was at the core of typography in the West until the twentieth century.

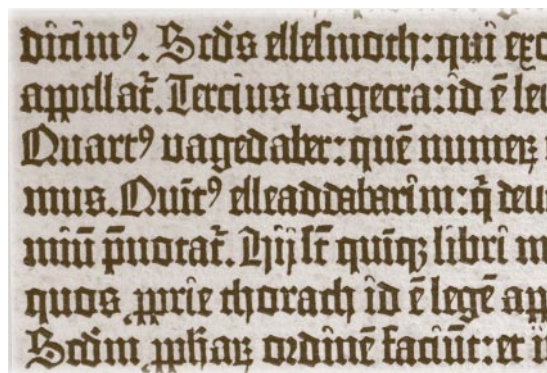
The reproducibility of printing types opens intriguing opportunities for the historian as archaeologist: beneath the abstraction of its textual content, a book can also be studied as a physical artefact. Watermarks and bindings provide valuable information about printer, provenance and

production, and recent advances in digital imaging allow detailed analysis of the printed characters themselves. We can begin to reconstruct the typefaces used in a printshop, trace the patterns of damage and wear to individual types, and study the early development of printing technologies. Were the changes in the look of typographic printing between 1450 and 1500 accompanied by corresponding technological changes? Only the surviving printed documents can answer; early printers left no account of their methods.

Printing drove the simplification of written language. Increasing literacy and the globalization of reading eventually led to standardization in spelling and punctuation, but a more fundamental change in the written representation of language occurred very rapidly in response to the new economics of print. By 1470, the Venetian printer Nicolas Jenson had developed a humanist Roman type, the first font specifically designed for print; by the sixteenth century, printed use of elaborate scribal shorthands was almost obsolete. Unlike modern font designers, however, Gutenberg and other early European printers sought to reproduce the handwritten look of contemporary manuscript. In fact, the Lambeth Palace copy of Gutenberg's famous 42-line Bible was misclassified as manuscript for 200 years.

The initial challenge in the creation of moveable type thus lay less in the aesthetics of the final product than in the segmentation of the written page into reusable pieces. In 1988, Janet Ing Freeman and Paul Needham, of the Scheide Library in Princeton, hand-compiled a set of 204 possible punches used by Gutenberg in his first font. Their source was a 1456 Bull of Pope Calixtus III, announcing the Turkish conquest of Constantinople. About 20% of the identified letterforms were repeats; though slightly different in shape, they had no distinct typographic meaning. This finding posed a minor mystery. Carving a steel punch requires a full day of work by a highly skilled craftsman. Each punch can strike many matrices; each matrix can yield a nearly limitless supply of type. Why would Gutenberg have carved so many unnecessary punches?

Computational analysis of the Calixtus Bull has helped to resolve this puzzle. Clustering all occurrences of a single letter allows



Sed postea quom maior orbis solaris factus esset: ignem in ipso fuisse intrusum. Epicurus Neoclis atheniensis filius maiestati deorum detrahere conatur: nec aliquid ex non ente fieri asserit: ac semper uniuersum fuisse habuisse neque aliquid noui fieri: praeter id quod tempore infinito in factum est: corpusque esse uniuersum non solum immutabile: uerum et infinitum finem bonorum uoluptatem statuit. Aristippus cyrenaicus uoluptatem bonorum: dolorem malorum finem constituit. Ceteras scientias excludit illud solum esse utile putans: ut queras si quid domi malum

The Gutenberg Bible (top), printed in 1454–55 and Nicolas Jenson's *Eusebius* of 1470, the first book to show off his humanist Roman type.

us to track type pieces as they are reused throughout the document. Although inking, paper grain and other factors add noise, the impressions fall into distinct clusters, each corresponding to a different type piece. Surprisingly, there are hundreds of versions of each letter. This suggests that Gutenberg did not make his types by the usual mass-production method using punches and matrices. Rather, they seem to have been handmade, one (or, at most, a few) at a time. Similar variability is observed in other Gutenberg printing, and in the work of several of his contemporaries. The inventor of the font is still unknown.

Thus, the Promethean view of the invention of typography as a single great leap must be slightly amended. The transition from manuscript to print was a two-way negotiation between the new technologies of page and type reproduction and the written representation of language itself. Gutenberg began this process by transmuting manuscript into moveable type, with magnificent results. Evolving selection pressures posed new problems for his successors. They needed to reverse the trend of the Gothic, moving towards lean, discrete and pan-European character sets: complex fonts are both costly to manufacture and difficult to typeset. This rebirth of classical simplicity in script had far-reaching repercussions, addressing in advance the challenge of our own transition to purely digital representations of text. ■

Blaise Agüera y Arcas is at the Scheide Library, Princeton University, Princeton, New Jersey 08544, USA. Adrienne Fairhall is at the NEC Research Institute, 4 Independence Way, Princeton, New Jersey 08540-6634, USA.

Predicting the future

John F. Y. Brookfield

The idea of fitness is central to evolutionary biology. The British philosopher Herbert Spencer characterized Darwin's theory as "the survival of the fittest" — the survival of the individuals that best fit their environment. In the 1930s, J. B. S. Haldane, Sewall Wright and Ronald Fisher quantified 'fitness' to express the strength of natural selection. Thus, if a mutant genotype suffers a 10% selective disadvantage relative to the wild type, it has a fitness of 90%.

This concept of fitness has prompted sterile debate along the lines that, as natural selection states that the fittest survive, and as 'the fittest' is defined as those that survive, the whole concept of natural selection is tautological. This misses the point: the main thing is that it is extremely useful to have a quantifiable, measurable description of a genotype.

Fitnesses of genotypes tend, empirically, to be roughly constant. Calculations based on this assumption give good predictions of the time course of the spread of advantageous alleles in populations. Fisher showed that when the fitness of each genotype is constant with time, the mean fitness of a population increases. Fitness describes the present success of a genotype, not the probability of its survival in the long term, which might depend on its capacity to adapt to new environments.

One fundamental misunderstanding is that any differences in survival or reproduction between individuals reflect differences in fitness. This is not what fitness means. Fitness represents an expected outcome, and what

actually happens in small populations differs from expectation because each generation's genotypes represent a sample, with an attendant sampling error, of the gametes produced by the previous generation; this is the basis of the phenomenon known as 'genetic drift'. The fitness of a genotype is related only probabilistically to real events; weakly advantageous mutations are usually lost by chance. Weakly deleterious mutations are much less likely to spread than advantageous alleles, but may arise much more often. Indeed, it is probable that most evolution of amino-acid sequences has occurred by fixation of weakly deleterious changes.

Fitness is hard to measure, particularly in wild populations, because it summarizes expected survival and reproduction. In particular, measuring 'lifetime reproductive success' by painstakingly tracking cohorts of individuals throughout their lives gives data that are difficult to interpret. Each individual in a sexual wild population has a genotype that, as an entity, is unique. An individual's genotype will have a fitness, which will thus be the individual's fitness. But random events cause the lives of individuals to differ, even those with identical fitnesses, and variation in the lifetime reproductive success of individuals does not represent variation in fitness between their genotypes. The only way to measure differences in fitness is to measure differences in mean survivorship and in mean reproductive rate between classes of individuals (groups of individuals that differ biologically). If one is interested in evolution, the only interesting classes are those that differ in their genotypes. A genotypic class, for example, might include all individuals that share the same genotype at a single genetic locus.

Now, if one is looking at the mean survival and reproduction of genetically defined classes of individuals, there is no point in looking at lifetime reproductive success. It is not worth measuring the fertility of young adults, for example, and then monitoring the fertility of the same individuals year after year as they grow older. You can find out all you need to know by looking at the fertility of different age classes in the same year — the fact that the individuals are different has no effect on the expected mean fitness.

Many view natural selection as an environmental force that acts on the phenotypes of populations, by analogy with artificial selection in animal or plant breeding. Although differences in genotypic fitness are caused by differences in phenotype, the widespread occurrence of pleiotropy — whereby a single genetic change has multiple phenotypic effects — means that it is very difficult to identify the true ways in which fitness differ-

Fitness

Fitness summarizes, rather than explains, the ability of a genotype to survive and reproduce.

ences arise. The most obvious phenotypic differences may not be the most important.

Does Fisher's theorem predict whether organisms become better adapted to their environment with time? In principle, yes, but there are important caveats. First, environments change. It is futile to try to explain human behaviour in adaptive terms, as the environments in which the genes responsible evolved were so different. For example, it may be that bad drivers crash cars more often, and so there is natural selection against individuals who are poor at driving. But this has no causal connection with the fact that humans can drive cars. The genes for this skill were not created by selection against individuals who crashed. Of course, nobody would suggest that they were, yet there are consistent, misguided attempts to explain other aspects of contemporary human behaviour in terms of the consequences, in effects on fitness, of those behaviours in modern environments.

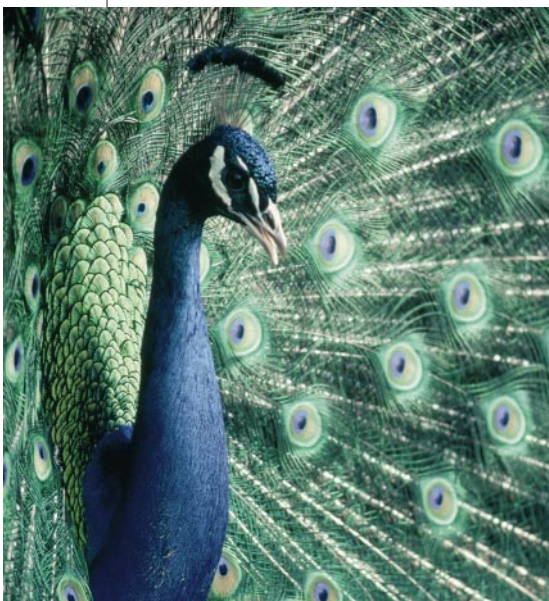
Second, although genes that improve survival tend to increase fitness, so do genes that increase sexual attractiveness — such as those that create the peacock's tail. A population of peacocks that did not evolve a spectacular tail might have been more successful in terms of population density or the probability of survival. Equally, in a population in which random genetic changes reduced male fitness to make all males slightly less attractive to females, the females would settle for second-best, and the species would get along fine. There is no necessary agreement between mean fitness and ecological variables.

I have used the term 'fitness' to describe differences between genotypes within species. What about the 'fitnesses' of different entities? Some species spread at the expense of others, and some ideas (memes) become better known by imitation. Can the spread of religion, for example, be explained in terms of 'memes' with high 'fitness', as some believe? Logically, 'fitness' could be used for these other entities, in which case its use would remain as circular and non-explanatory as it is for genotypes. But here, fitness is not constant over time, so there is no pragmatic justification for it as a predictive mathematical tool. ■

John F. Y. Brookfield is at the Institute of Genetics, University of Nottingham, Nottingham NG7 2UH, UK.

FURTHER READING

Smith, J. M. *Evolutionary Genetics* 2nd edn (Oxford Univ. Press, Oxford, 1998).



Show-off: genes to woo females can boost fitness.

Are all cancer genes equal?

Jiri Bartek and Jiri Lukas

Cancer arises because of the accumulation of defects in certain classes of genes. In mice, breast tumours caused by one such class of 'cancer genes' can be prevented.

When they become cancerous, mammalian cells tend to either lose or 'silence' tumour-suppressor genes — the protein products of which, as the name implies, normally help to suppress cancerous features of a cell. The cells also mutate or overexpress certain other genes, whose products in normal circumstances promote, for example, healthy cell proliferation or cell survival. Mutation or overexpression turns these genes, known as proto-oncogenes, into the hyperactive oncogene form. These genetic abnormalities tend to occur in combinations that vary from tissue to tissue — a fact that has long puzzled cancer researchers. The patterns might relate to the various ways in which different cells control their multiplication and specialization, and a better understanding of the tissue-specific effects may help in treating particular tumours.

On page 1017 of this issue¹, Yu and colleagues provide stunning evidence in support of this idea, based on their studies of mice that are genetically modified to be prone to breast cancer. Their unexpected results show that mice lacking cyclin D1, one of many proteins involved in cell proliferation, are entirely resistant to breast tumours induced by a class of oncogenes that have been implicated in most human breast cancers.

Cyclins are proteins that are essential for cell division. They work by activating their partners, the cyclin-dependent kinases, and directing these enzymes to specific substrates². One of the subgroups of cyclins consists of the three known 'D-type' cyclins, D1, D2 and D3, which accelerate the G1 phase of the mammalian cell-division cycle — a period during which a cell makes fundamental decisions, such as whether to proliferate or become specialized in some way². Like their proto-oncogenic counterparts, many oncogenes stimulate cell proliferation by inducing the expression of the D-type cyclins, particularly cyclin D1 (refs 2, 3).

The genes encoding D-cyclins are themselves proto-oncogenes; they can be turned into oncogenes by viruses or by overexpression in several types of tumour, including — in the case of cyclin D1 — some 50% of human breast cancers¹⁻⁴. Cyclin D1 is also required for the proliferation of breast-cancer cells in culture⁴. The link between cyclin D1 and the proliferation of breast

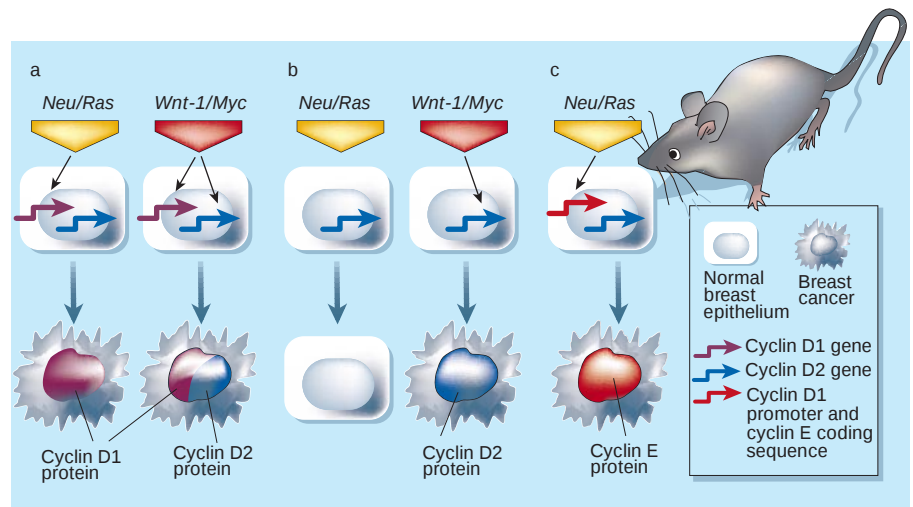


Figure 1 Mice lacking the cyclin D1 gene are resistant to some types of breast cancer, as shown by Yu *et al.*¹. **a**, In normal breast epithelium, in which both the cyclin D1 and the cyclin D2 genes are intact, exposure to various classes of oncogene promotes cancer formation. This is accompanied by activation of the cyclin D1 (but not D2) gene by the Neu and Ras oncogenes, and of both cyclin D1 and D2 by the Wnt-1 and Myc oncogenes. **b**, When the cyclin D1 gene is disrupted, Neu and Ras can no longer stimulate cell proliferation or breast tumours through cyclin D. But Wnt-1 and Myc, which promote expression of cyclin D2, can still induce tumours in cyclin-D1-deficient breast epithelium. **c**, When the cyclin D1 promoter (a regulatory region) is fused to the protein-coding part of the cyclin E gene, Neu/Ras can induce breast cancer. So, Neu and Ras signal through elements within the cyclin D1 promoter. As Wnt-1 activates Myc, and Neu activates Ras, we show these oncogenes as pairs (Wnt-1/Myc and Neu/Ras).

epithelial cells (which line the ducts and alveoli of the sponge-like mammary glands) is perhaps best highlighted by the specific consequences of genetically removing cyclin D1 in mice^{5,6}. This change has almost no effects on the physiology of adult female mice, other than impairing breast development during pregnancy^{5,6}.

All of this inspired Yu *et al.*¹ to test whether the removal of cyclin D1 could prevent breast cancer without harming other tissues. They did this by crossing cyclin-D1-deficient mice with four other groups of mice, each modified to express a different oncogene known to cause breast tumours. The authors found that the absence of cyclin D1 did not protect mice from developing breast cancers induced by two of these oncogenes, called Wnt-1 and Myc. In sharp contrast, the cyclin-D1-deficient mice were entirely resistant to breast cancers induced by the other two oncogenes studied, Neu and Ras, which evoked numerous breast tumours in control mice (Fig. 1a, b). Mice deficient in cyclin D2 or cyclin D3 remained susceptible to all four oncogenes, indicating that the dramatic protection against Neu-

or Ras-induced breast tumours resulted only from lack of cyclin D1.

While searching for an explanation of such clear-cut differences between the two groups of oncogenes (Wnt-1 and Myc versus Neu and Ras), Yu *et al.*¹ noticed that, in normal mice, Myc and Wnt-1 increased the expression of both cyclin D1 and cyclin D2 in breast tumours. But Neu and Ras led to high levels of expression only of cyclin D1 in these cancers. The authors went on to show that the strict requirement for cyclin D1 in Neu/Ras-induced breast cancer does not apply to other cell types, such as salivary-gland cells. The implication is that, in breast epithelium, Neu and Ras promote cell proliferation only by inducing the expression of cyclin D1. Wnt-1 and Myc, meanwhile, can work through other targets, including cyclin D2. In tissues other than breast, Ras and Neu can also work through other D-type cyclins.

Moreover, the authors show that Ras and Neu work specifically to regulate the promoter region of the cyclin D1 gene — a region that does not encode the protein product but is needed to control gene expression. They discovered this by replacing the

coding sequences (but not the promoter) of the cyclin D1 gene with those of another type of cyclin, cyclin E (ref. 7), in female mice; the incidence of breast cancer was similar to that in mice with the intact cyclin D1 gene (Fig. 1c). The next step is to work out the unique cellular events that are induced by *Neu* and *Ras* and which in turn control the cyclin D1 promoter specifically in breast epithelium.

Such tissue-dependent mechanisms might not be restricted to cyclin D1 and breast cancer. Mice lacking cyclin D2 suffer defects specifically in their testes or ovaries, and overabundant cyclin D2 has been implicated in tumours arising from these organs⁸. Moreover, it could be that mice lacking cyclin D3 have defects in their blood cells or lymphoid organs (or both), as cyclin D3 is the main D-type cyclin in some blood-cell lineages and in some types of leukaemia and lymphoma. More broadly, the work of Yu *et al.*¹ may encourage researchers to demystify other tissue-specific phenomena; there are many gene families whose members are expressed in a tissue-dependent way.

Yu *et al.*¹ conclude that human breast cancers that display activated *Neu* or *Ras* could be prime targets for future drugs aimed at cyclin D1. Such drugs might either bind to cyclin D1, blocking its function, or decrease the level of this cyclin in cancer cells. This optimism might be justified, as large subsets of human breast tumours have hyperactive *Neu* or *Ras* and an overabundance of cyclin D1 but not D2 (refs 3, 4, 9). But both the molecular defects and the

clinical features vary considerably among even these breast-cancer patients. And there are other issues to investigate, such as the involvement of oestrogen receptors, which are abundant in breast tumours, and the seemingly paradoxical silencing of the cyclin D2 promoter in subsets of human breast cancers⁹. New breast-cancer treatments that target the *Neu* oncogene itself³ are in clinical trials, and it remains to be seen whether drugs aimed at cyclin D1 (if such drugs are developed), or a combination of drugs, might offer any benefits or minimize side effects in other tissues. If so, it might one day be possible to provide such tailor-made treatments, selecting appropriate breast-cancer patients by molecular profiling using DNA chips, or by assays to study cyclin D1, *Ras* and *Myc*, for example.

All oncogenes can deregulate cell growth. But it is clear that, to paraphrase George Orwell, some oncogenes are more equal than others. Let us hope that we can turn this knowledge to our advantage. ■

Jiri Bartek and Jiri Lukas are in the Department of Cell Cycle and Cancer, Institute of Cancer Biology, Danish Cancer Society, Strandboulevarden 49, DK-2100 Copenhagen, Denmark.

e-mail: bartek@biobase.dk

1. Yu, Q., Geng, Y. & Sicinski, P. *Nature* **411**, 1017–1021 (2001).
2. Sherr, C. J. & Roberts, J. M. *Genes Dev.* **13**, 1501–1512 (1999).
3. Harari, D. & Yarden, Y. *Oncogene* **19**, 6102–6114 (2000).
4. Bartkova, J. *et al.* *Int. J. Cancer* **57**, 353–361 (1994).
5. Sicinski, P. *et al.* *Cell* **82**, 621–630 (1995).
6. Fantl, V., Stamp, G., Andrews, A., Rosewell, I. & Dickson, C. *Genes Dev.* **9**, 2364–2372 (1995).
7. Geng, Y. *et al.* *Cell* **97**, 767–777 (1999).
8. Sicinski, P. *et al.* *Nature* **384**, 470–474 (1996).
9. Evron, E. *et al.* *Cancer Res.* **61**, 2782–2787 (2001).

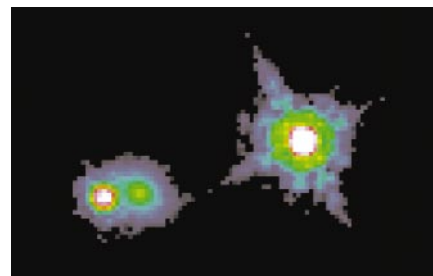


Figure 1 A double image of a distant quasar, Q0143, caused by a massive galaxy (the 'lens') located close to the line of sight between the quasar and the observer. The gravitational lens is apparently close to the fainter image (left). This is one of many examples of gravitational lensing from the CASTLE collaboration (see <http://cfa-www.harvard.edu/glensdata/image.html>). When the gravitational lens is much smaller — say, a single star or planet — the double images are too close to be seen separately. But predictable changes in their brightness are taken as evidence for 'gravitational microlensing'. Sahu *et al.*³ have discovered microlensing events produced by individual objects in a globular cluster that may be small enough to be free-roaming planets.

Sahu *et al.*³ report the first ever detections of microlensing events for which they can determine the lens masses. If some of these objects are as light as they appear to be, they would be free-roaming planets.

The probability of a microlensing event occurring in our Galaxy, or in a neighbouring one, is very small — one part in a million, or less. So to detect intensity variations due to gravitational microlensing, millions of stars have to be monitored. The hunt started decades ago, and several hundred events have since been discovered, with durations ranging from a few days to many months, more or less as expected for stellar-mass lenses. Some events were truly spectacular, with the apparent brightness of the star increasing 100-fold. Most searches were aimed at finding evidence for the Universe's missing 'dark matter', because the effect depends on the mass of the lens, not on its brightness. Unfortunately, most lensing objects appear to be stars, so the dark-matter hunters still haven't found what they are looking for.

As well as being hard to detect, gravitational microlensing suffers from a major drawback. The mass of the lens is proportional to the square of the duration of the microlensing event, but it is also affected by the distance and transverse velocity of the lens. These are generally not known, and can only be guessed at statistically. So although the durations of the observed events are consistent with the expectations of stellar-mass lenses, we do not know the specific masses of the lenses discovered so far.

This is where Sahu *et al.*³ come in. They picked a member of a globular star cluster as their lens⁴. We know the distance and velocity

Astronomy

Planetary candidates

Bohdan Paczynski

The tentative discovery of planets roaming freely through interstellar space has far-reaching implications. If confirmed, it would imply that there are more planets on the loose than there are around stars.

A solar eclipse in 1919 provided the first evidence that a massive object, in this case the Sun, can bend light from distant stars, making them appear to be in the wrong position. In subsequent decades, astronomers started trying to observe the 'gravitational lensing' effect — this occurs when the gravitational field of a massive object (the lens), located close to the observers' line of sight, bends the light from a distant object, making it appear to be in two different places a small distance apart (Fig. 1). The first case² was discovered in 1979, and several dozen double and even quadruple images of distant quasars and galaxies have since been found, with galaxies, or clusters of galaxies, acting as the gravitational lenses.

Individual stars and planets, although far

less massive than galaxies, could in principle also act as gravitational lenses. However, the predicted separation between the two images created by these lenses is too small to be resolved by any optical telescope. Nonetheless, as the source (a distant star), the lens (a star or planet close to the line of sight), and the observer move with respect to each other, the geometry of lensing changes. The intensity of the two images changes accordingly, and their combined brightness varies in a predictable way. As the two images cannot be seen separately, this phenomenon is called gravitational microlensing. These changes in brightness typically take place over a period of a few weeks for lenses of stellar mass, and hours or days for lenses the size of planets. On page 1022 of this issue,

of most globular clusters. There are about one hundred such clusters in our Galaxy, each of which contains about 100,000 stars. The stars within a cluster move randomly, with velocities of several kilometres per second with respect to the cluster centre, while the cluster as a whole moves at up to 200 kilometres per second with respect to the centre of our Galaxy. Some of the clusters are seen against a rich background of distant stars. Monitoring those distant stars for variability may reveal rare cases of gravitational microlensing caused by stars and planets located within the globular clusters.

How do Sahu *et al.* know that a particular microlensing event is caused by an object in a star cluster? They know simply because the mass concentration near the centre of the cluster is so high that the probability of microlensing by a cluster member is far higher than by a random star located somewhere else along the line of sight. However, there are so many stars near the centre of the cluster that it is hard to see those in the background. This is why Sahu *et al.* used the Hubble Space Telescope — its images are at least ten times sharper than those obtained with any ground-based telescope.

Sahu *et al.*³ monitored the brightness of about 83,000 stars located behind a globular cluster known as M22. The results were both predictable and surprising. The brightness of one of the stars increased more than ten-fold, over a period of 17.6 days, corresponding to a lens of 0.13 solar masses — about the size of a faint red dwarf star. The surprising thing about this discovery is that the lens appears to be a rare type of variable star, which has small periodic variations in light amplitude that affect the lensing in a predictable way. Only a small fraction of stars vary like this, and it is unexpected to find a rare star in the first instance.

Sahu and colleagues also observed that six other background stars briefly increased in brightness, typically by 50%. The duration of these events, if confirmed as gravitational microlensing, was less than a day, suggesting a lens smaller in mass than Jupiter. These planetary-mass events are harder to interpret than the stellar lens, because they are based on a single measurement, rather than a light curve. A light curve consists of several dozen photometric measurements whose shape can reveal the underlying cause of variability. If these six objects are accepted as gravitational microlensing, then about 10% of the total mass of M22 comes from objects with masses similar to Jupiter or Saturn, but not associated with any star — planets on the loose.

What can we make of these discoveries? In my view, the situation is similar to that of 1993 when three collaborations reported the first candidate microlensing events^{5–7}. Today — several hundred detections later — there is plenty of evidence that the stellar variability

they detected is due to gravitational microlensing, and the events are no longer considered as mere 'candidates'. I consider the discoveries of Sahu *et al.* to be at the level of the 1993 events because gravitational microlensing is not the only possible interpretation.

But these discoveries do justify further searches. The stellar-mass event should be looked at again for new activity, and for the candidate planetary events we need much more frequent measurements — at least once every 90 minutes (the duration of the Hubble telescope's orbit), and preferably several times every 90 minutes, to obtain full light curves. Only when future observers establish that the shapes of the light curves agree with predictions of gravitational microlensing can we be confident that a

significant fraction of the M22 mass is made of planetary objects. If this is done and the interpretation of Sahu *et al.* is confirmed, we shall have a truly spectacular discovery. For now, we have a fascinating result, possibly a preview of coming attractions. ■

Bohdan Paczynski is in the Department of Astrophysical Sciences, Princeton University Observatory, Princeton, New Jersey 08544-1001, USA.

e-mail: bp@astro.princeton.edu

1. Dyson, F. W., Eddington, A. S. & Davidson, C. *Phil. Trans. R. Soc. Lond. A* **220**, 291 (1920).
2. Walsh, D., Carswell, R. F. & Weymann, R. J. *Nature* **279**, 381 (1979).
3. Sahu, K. C. *et al. Nature* **411**, 1022–1024 (2001).
4. Paczynski, B. *Acta Astron.* **44**, 235–239 (1994).
5. Alcock, C. *et al. Nature* **365**, 621–623 (1993).
6. Aubourg, E. *et al. Nature* **365**, 623–625 (1993).
7. Udalski, A. *et al. Acta Astron.* **43**, 289–295 (1993).

Biological catalysis

Repopulating the RNA world

Scott A. Strobel

We will probably never be certain how life on our planet began. But, thanks to a new laboratory study, the possibility of an ancient 'RNA world' appears more plausible.

In life as we know it, the tasks of storing genetic information and of copying it for transfer to subsequent generations are clearly divided. Nucleic acids encode the information, and proteins replicate it. But, during the early evolution of life on Earth, it seems likely that these tasks were shared by one type of molecule — and the leading contender is RNA. A key requirement of this 'RNA world' is the existence of an RNA-copying enzyme that is itself composed of RNA¹. No such molecular fossil has yet been found, but Bartel and colleagues² report in

Science that they have managed to develop an RNA with this type of activity.

The central dogma of molecular biology, as stated in simplistic terms and found in every biochemical textbook, reads: 'DNA encodes RNA and RNA encodes protein'. But the original formulation of the central dogma as presented by Francis Crick³ was not explicitly a biochemical hypothesis. Instead it was a prediction about the transfer of cellular information. "This [the central dogma] states that once 'information' has passed into protein it cannot get out again. In

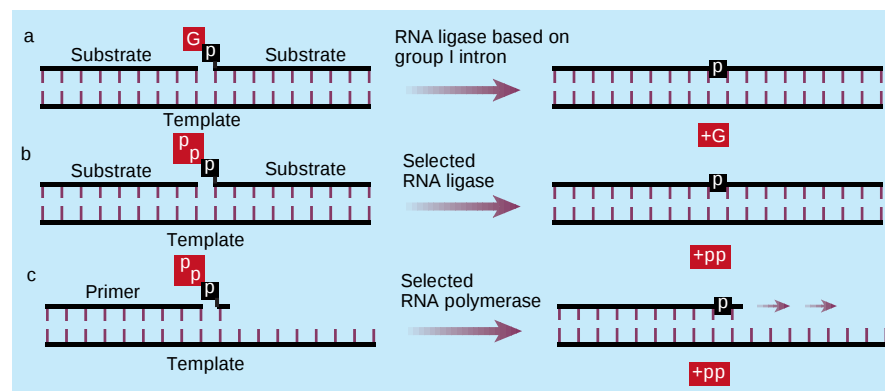


Figure 1 The search for an RNA-based enzyme (ribozyme) that replicates RNA. **a**, Experimental modifications to a self-splicing RNA molecule known as a group I intron led to the production of a ribozyme that can join (ligate) two RNA fragments (substrates) aligned on a template, using guanosine (G) as the leaving group⁷. **b**, Another ribozyme can perform a similar ligation reaction, but using the more typical pyrophosphate (pp) as the leaving group⁸. **c**, Bartel and colleagues² started with the core ligase activity of the enzyme from **b**, and forced it to 'evolve'. From the resulting RNAs, the authors identified an RNA-copying ribozyme (RNA polymerase) that can use single nucleotide triphosphates as a substrate. This enzyme adds up to 14 nucleotides to a 'primer' sequence that is complementary to the RNA template. 'p' represents phosphate.

more detail, the transfer of information from nucleic acid to nucleic acid, or from nucleic acid to protein may be possible, but transfer from protein to protein, or from protein to nucleic acid is impossible." By 'information', Crick meant a precise sequence, either of bases in a nucleic acid or of amino-acid residues in a protein. So, a fundamental tenet of molecular biology is that a protein constitutes an informational dead-end to the cell.

Put another way, a protein is unable to replicate itself, not because it lacks catalytic capabilities but because its information is irretrievable. So, one of the main aims of those investigating prebiotic molecular evolution is to discover (or re-invent) a molecule with the catalytic prowess of a protein, but the accessible information content of a nucleic acid¹.

For several reasons, RNA is the target of this search. RNA polymerization — which involves the reading of a DNA or RNA template and production of a complementary RNA strand — is likely to have been the responsibility of protein-based enzymes since cellular life began. But enough catalytic RNAs exist to infer that RNA might once have been its own replicating enzyme. Natural catalytic RNAs (ribozymes) have been discovered that enhance reactions similar to those required for replication⁴. Furthermore, ribosomal RNA acts as the catalyst of

protein synthesis, so polymerase activity *per se* (in this case, protein polymerization) is not beyond the scope of RNA⁵. Nevertheless, a possible 'founder' polymerase is either long extinct or as yet undiscovered, leaving biochemists to conjure up proof-of-principle schemes for repopulating a presumptive RNA world.

The reconstruction of an RNA-replicating ribozyme began with experimental modifications of a naturally occurring, self-splicing RNA molecule known as a group I intron⁶. The modified molecules could covalently join several RNA sequences, aligned along an RNA template, efficiently enough to generate a few full-length RNA strands up to 200 nucleotides long⁷ (Fig. 1a). Unfortunately, the RNA sequences to be joined had to be about ten nucleotides in size before any appreciable activity was seen, and a 5'-terminal guanosine, instead of the more common pyrophosphate, was the leaving group in the joining reaction. A true RNA polymerase, as we would now define it, must be able to use single nucleotides and join them accurately in the order specified by an RNA template.

The next step came in the form of RNAs that could join two RNA fragments aligned along an RNA template; but they use pyrophosphate as the leaving group⁸ (Fig. 1b). These RNAs are still 'ligases' — they do not copy RNA but merely join two RNA

molecules — but they promote the same chemical joining reaction as that performed by protein-based polymerases. The RNA enzymes here⁸ were obtained not from the group I intron, but by *in vitro* amplification and darwinian selection from an enormous pool of random RNA sequences, each of which was more than 100 nucleotides long. Among the active sequences selected from this morass was one with an unusually complex knot-like structure, which had particularly robust efficiency⁹. This core ligase provided Bartel and co-workers with the fundamental catalytic capability needed to identify a simple polymerase¹⁰.

Starting from the core ligase⁹, and performing several iterations of mutation and selection, Bartel and colleagues² have developed a new variant that comes even closer to the properties expected of a true RNA-templated, RNA-catalysed, RNA polymerase. Unlike its predecessors, this polymerase has no restrictions on the sequence it replicates. It can synthesize more than a full turn of an RNA helix (up to 14 nucleotides), adding single nucleotide triphosphates to a short primer sequence that is complementary to the RNA template (Fig. 1c). The polymerase can distinguish correct from incorrect primers, and it preferentially extends matched primers with an accuracy of 96–99%, a fidelity on a par with some protein-based polymerases¹¹.

Yet this polymerase still lacks at least two properties expected of the prebiotic RNA-replicating enzyme: it cannot copy longer templates fully, and its polymerization is not efficient enough to produce progeny RNA molecules at a rate exceeding the rate of decomposition of parental RNAs. Furthermore, given the slow rates of template copying, the polymerase does not appear to be 'processive' — after adding the first nucleotide, the polymerase is more likely to dissociate from the template than to remain associated and add a second nucleotide. Self-complementary regions of the template could form structures that block access by the polymerase. A more processive polymerase, or an enzyme with a companion 'unwinding' activity, may be needed to achieve efficient polymerization. In fact, an unwinding activity might be particularly important because self-replication is actually a two-step process: synthesis of the first strand results in an extended duplex, which must be peeled apart before the new strand can be copied to regenerate the original sequence.

Bartel and colleagues' latest polymerase is relatively large (189 nucleotides) and structurally complex, so the question remains as to how such an RNA could have evolved in the first place. This becomes a matter of the wildest speculation, but the general path from ligase to polymerase used in the *in vitro* selection strategy² might provide clues. The

Geophysics

Martian motion

At some time in the past, an object hit Mars near its south pole. The result was a crater, 45 km in diameter, part of which is evident as the green crescent at the left (west) of this image compiled from Mars Orbiter Laser Altimeter data. But why a crescent? As he describes in the *Journal of Geophysical Research* (**106**, 10075–10085; 2001), James W. Head concludes that the crescent shape resulted from considerable movement of the martian polar cap within the past few million years. Head

constructed profiles of the crater and neighbouring areas, and took especial account of the pattern of an 'ejecta lobe' and secondary ejecta craters seen on higher-resolution pictures.

The south pole of Mars is thought to be covered by seasonally shifting deposits of water-ice, usually with a layer of CO₂ on top. These deposits overlie a much more stable 'polar layered terrain' of ice and dust. Head's thesis is that, in the comparatively recent past, polar layered terrain swept from the

south of the crater to occupy a large part of it, leaving only the original crater floor, seen here in green. This conclusion is largely based on the pattern of surviving secondary craters — those on the ground underlying the layered terrain (brown) close to the impact, or in the path of the part of the polar cap that moved, being obliterated. From features known as mantled deposits — residues of polar layered terrain activity — to the north and east of the crater, Head also surmises that the layered terrain later retreated partially.

The other large impact crater, seen on the right of this picture, lies 300 km from that painstakingly investigated by Head. Its western part likewise contains a lobe of polar layered terrain, a further indication of the possible occurrence of dynamic processes at the martian south pole. **Tim Lincoln**



authors argue that the polymerase would have been much more difficult, if not impossible, to identify without using the core ligase activity as a starting point. Others have speculated that darwinian evolution began even before the RNA world, with an unidentified RNA-like polymer that also had catalytic features and could use a template, but had a more accessible synthesis under prebiotic conditions¹. Although the evolutionary events that might have led to an RNA world are not known and are potentially unknowable, Bartel and colleagues' results² lend further credibility to the RNA-world hypothesis for the early evolution of life. ■

Scott A. Strobel is in the Department of Molecular

*Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520-8114, USA.
e-mail: strobel@csb.yale.edu*

1. Joyce, G. F. & Orgel, L. E. in *The RNA World* (eds Gesteland, R. F. & Atkins, J. F.) 1–25 (Cold Spring Harbor Lab. Press, New York, 1993).
2. Johnston, W. K., Unrau, P. J., Lawrence, M. S., Glasner, M. E. & Bartel, D. P. *Science* **292**, 1319–1325 (2001).
3. Crick, F. H. C. *Symp. Soc. Exp. Biol.* **12**, 138–163 (1957).
4. Cech, T. R. *Science* **236**, 1532–1539 (1987).
5. Nissen, P., Hansen, J., Ban, N., Moore, P. & Steitz, T. *Science* **289**, 920–930 (2000).
6. Doudna, J. A. & Szostak, J. W. *Nature* **339**, 519–522 (1989).
7. Green, R. & Szostak, J. W. *Science* **258**, 1910–1915 (1992).
8. Bartel, D. P. & Szostak, J. W. *Science* **261**, 1411–1417 (1993).
9. Eklund, E. H., Szostak, J. W. & Bartel, D. P. *Science* **269**, 364–370 (1995).
10. Eklund, E. H. & Bartel, D. P. *Nature* **382**, 373–376 (1996).
11. Ward, C. D. & Flanagan, J. B. *J. Virol.* **66**, 3784–3793 (1992).

Cell biology

Glowing switches

Johannes L. Bos

An approach involving changes in the emission of fluorescence reveals new information about the activation of two important molecular switches in real time and space.

Many proteins function as simple on–off switches. Take, for example, the proteins of the so-called Ras family, which are molecular switches that regulate a variety of cellular processes, including the pathways by which information is received at a cell's surface and passed on inside the cell. Ras proteins are in the 'on' form when bound to guanine nucleotide triphosphate (GTP), and are switched off when bound to guanine nucleotide diphosphate (GDP). Until now, the activation of these proteins has been determined biochemically. But, on page 1065 of this issue¹, Mochizuki and colleagues describe how they used the technique of fluorescent resonance energy transfer (FRET) to monitor the activation of two such proteins in living cells. For the first time, we can see that the eponymous member of the family, Ras itself, is switched on at locations near the cell membrane. Intriguingly, however, a close relative — Rap1 — is activated near the nucleus.

Proteins from the Ras family are anchored to the cytoplasmic sides of cellular membranes through a covalently bound lipid group. In general, they are switched on in response to the activation of certain receptor molecules, which detect changes — such as the presence of growth factors — outside the cell. Inside the cell, the receptors are often connected to guanine-nucleotide-exchange factors, which replace GDP with GTP in Ras proteins. The Ras proteins are thereby switched on, and in turn influence the activity of 'effector' molecules. Ras proteins switch themselves off by hydrolysing bound GTP to GDP, and are therefore known as small GTP-hydrolysing enzymes (GTPases).

GTPase-activating proteins greatly enhance this hydrolysis.

The activation of GTPases can be measured biochemically by using a region of an effector molecule that binds specifically to the active protein. For instance, the Ras-binding domain of the Raf1 protein interacts with GTP-bound, but not GDP-bound, Ras. Such interactions can also be monitored by FRET analysis, as first shown for another small GTPase, Rac². Mochizuki *et al.*¹ have now developed a new way of using FRET to analyse the activation of small GTPases.

The authors engineered a protein containing a GTPase (either Ras or Rap1) fused to yellow fluorescent protein, as well as the Ras-binding domain of Raf1 bound to cyan fluorescent protein (Fig. 1). They inserted this protein into different cell types, and used FRET to determine where and when the GTPase was activated. FRET basically involves the transfer of energy from one fluorescent molecule to another, which then emits its own fluorescence. It occurs only when the two fluorescent molecules are close to each other. In this case, a change in fluorescence occurs when the GTPase is in the active form and so interacts closely with the Ras-binding domain. Having the Ras-binding domain in the same molecule as the GTPase circumvents potential problems of the domain interfering with the function of the normal cellular GTPase and so affecting cell viability.

The authors go to great lengths to show the specificity of their fusion proteins. The Ras construct responded to Ras-specific activating and deactivating factors only, and the Rap1 construct responded only to Rap1-specific control factors. One possible drawback is that the constructs might not be found at the correct cellular site, even though they include membrane-anchoring signals. But, as the localization of the control factors themselves is also regulated by their interactions with other proteins or lipids, it is likely that the activation of the fluorescent GTPases reflects activation of the cellular proteins. Moreover, the positions at which the constructs are activated are consistent with previous studies of the localization of Ras and Rap1 (ref. 3).

The results of Mochizuki *et al.*¹ can best be appreciated by watching the videos

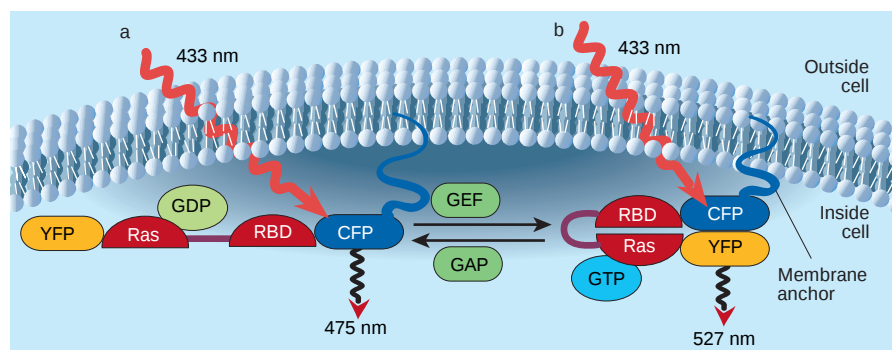


Figure 1 The technique used by Mochizuki *et al.*¹ to analyse the activation of Ras, a small GTPase enzyme, in real time and space. The authors first created a protein containing Ras, the Ras-binding domain (RBD) of the protein Raf1, and two fluorophores, yellow fluorescent protein (YFP) and cyan fluorescent protein (CFP). **a**, When Ras is inactive (in its GDP-bound form), the emission profile of CFP peaks at 475 nm when it is excited by light of wavelength 433 nm. **b**, When Ras is active (bound to GTP), the protein folds up so that the Ras-binding domain interacts with Ras. The two fluorophores are now in close proximity. The energy emitted by CFP is partly captured by YFP, which emits energy with a profile that peaks at 527 nm. The ratio of emission at 527 and 475 nm can be monitored in real time and space by using computer-enhanced time-lapse video microscopy, providing a picture of the pattern of activation of Ras. Guanine-nucleotide-exchange factors (GEFs) and GTPase-activating proteins (GAPs) activate or inactivate Ras, respectively. The authors¹ used a similar approach to study Rap1, another small GTPase.

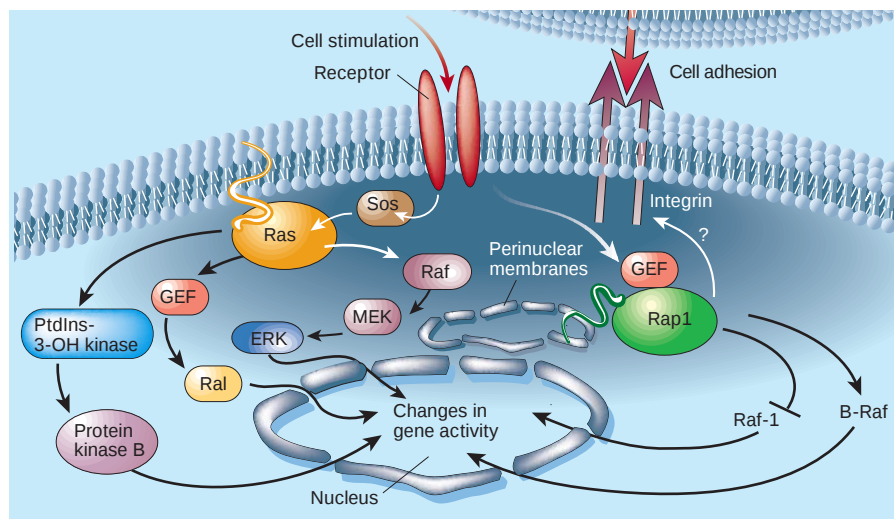


Figure 2 Ras and Rap1 are active in different parts of the cell. Mochizuki *et al.*¹ find that, in response to cell stimulation, Ras is active near the cell membrane, whereas Rap1 is switched on near the membranes around the nucleus. This may reflect the involvement of these proteins in signalling pathways that lead in opposite directions, either into (Ras) or out of (Rap1) the cell. The molecules involved in some of these pathways are shown. GEF, guanine-nucleotide-exchange factor; PtdIns-3-OH kinase, phosphatidylinositol-3-OH kinase.

provided as Supplementary Information (<http://www.nature.com/nature>). When the authors stimulated Cos cells with epidermal growth factor, they first detected activated Ras at the nearby plasma membrane. Active Ras then rapidly spread into the body of the cell, probably reflecting the internalization of a signalling complex composed of Ras, a guanine-nucleotide-exchange factor and perhaps the receptor. In PC12 cells that were induced to differentiate with nerve growth factor, active Ras was also detected near the plasma membrane, and persisted particularly at the tips of the extensions (neurites) sent out by the differentiated cells. This persistence was presumably due to continuous activation of Ras — the fluorescent signal was rapidly restored after being inactivated by photobleaching.

In contrast, the activation of Rap1 began just outside the nucleus and spread outwards. This implies that information from cell-surface receptors is transduced into the cell before Rap1 is activated. Indeed, the activation of Rap1 requires, for instance, the internalization of cell-surface receptors^{1,4}, or messenger molecules such as calcium that move freely in the cell (reviewed in ref. 5).

For Ras, these results are expected. Generally, a guanine-nucleotide-exchange factor, Sos, interacts (through connecting proteins) with receptors soon after they are activated. Sos then switches on Ras, which in turn activates several signalling pathways (Fig. 2), the end point of which is usually the control of gene expression³. The observations of Rap1, however, are more surprising. This protein is rapidly activated by several extracellular stimuli and, depending on the cell type, is implicated in either positive or negative control of a pathway that ultimately

regulates gene expression. Rap1 has also been proposed to coordinate some functions of the plasma membrane, such as cell–cell adhesion, but little is known about how it regulates these processes⁵.

The results of Mochizuki *et al.*¹ give a twist to our thinking about Rap1. It seems that the Rap1 signalling pathway responds to extracellular stimuli but begins inside the cell and directs signals outwards, for instance towards the plasma membrane. Why should Rap1 be activated near the nucleus to control events

at the plasma membrane? This protein is often found near the nucleus in organelles known as endosomes⁶, one of the functions of which is to sort proteins that are destined for different locations in the cell⁷. Rap1 may therefore be involved in protein sorting. It is also found in membranous packets of proteins that are destined to move outwards⁸, and may even be needed to regulate secretion, as suggested by the involvement of one of its exchange factors in this process⁹. Finally, Bud1, the budding-yeast counterpart of Rap1, is involved in assembling protein complexes at the place on the plasma membrane where a bud will form¹⁰.

The implication of all of these results is that Rap1 is involved in collecting together proteins for transport to the plasma membrane. I predict that Mochizuki *et al.*'s technique¹, together with methods for determining the precise membranes in which active GTPases are found, will be invaluable in tackling this and related problems. ■

Johannes L. Bos is in the Department of Physiological Chemistry, Utrecht University, Universiteitsweg 100, NL-3584 CG Utrecht, The Netherlands.

e-mail: j.l.bos@med.uu.nl

1. Mochizuki, N. *et al.* *Nature* **411**, 1065–1068 (2001).
2. Kraynov, V. S. *et al.* *Science* **290**, 333–337 (2000).
3. Bos, J. L. *EMBO J.* **17**, 6776–6782 (1998).
4. York, R. D. *et al.* *Mol. Cell. Biol.* **20**, 8069–8083 (2000).
5. Bos, J. L., De Rooij, J. & Reedquist, K. A. *Nature Rev. Mol. Cell Biol.* **2**, 369–377 (2001).
6. Pizon, V., Desjardins, M., Bucci, C., Parton, R. G. & Zerial, M. *J. Cell Sci.* **107**, 1661–1670 (1994).
7. Lemmon, S. K. & Traub, L. M. *Curr. Opin. Cell Biol.* **12**, 457–466 (2000).
8. Quinn, M. T. *et al.* *Blood* **79**, 1563–1573 (1992).
9. Ozaki, N. *et al.* *Nature Cell Biol.* **2**, 805–811 (2000).
10. Gulli, M.-P. & Peter, M. *Genes Dev.* **15**, 365–377 (2001).

Communications technology

A bottleneck for optical fibres

Joseph M. Kahn and Keang-Po Ho

Optical fibres carry light signals at extremely high frequencies, and offer enormous bandwidth for transmitting data. But nonlinear effects may limit their capacity to carry information.

Over the past decade, the flow of Internet traffic and other digital information has increased explosively. Optical fibres made of silica offer the highest capacity of any communications medium, making them the most cost-effective choice for data transmission over all but the shortest distances (under about 1 km). Even so, transmitting and receiving light signals also requires electronic systems that convert electricity into light and vice versa. These conversion processes tend to limit the rate at which information can be carried on an optical fibre using a single frequency or wavelength. But dramatic increases in the capacity of optical fibres have been achieved since the mid-1990s by sending numerous

signals along a single fibre, each with a different wavelength.

Commercial versions of this technology, called dense wavelength-division multiplexing (DWDM; Fig. 1, overleaf), achieve data transmission rates as high as 1.6 terabits per second by combining 160 signals that each carry 10 gigabits per second (Gbit s^{-1}). These systems transmit signals at infrared wavelengths between 1.53 and 1.60 μm . Optical amplifiers, placed at intervals of 60 to 100 km, simultaneously amplify signals at many wavelengths, allowing them to propagate over thousands of kilometres. On page 1027 of this issue, Mitra and Stark¹ calculate the fundamental information capacity of typical optical fibres, which is ultimately

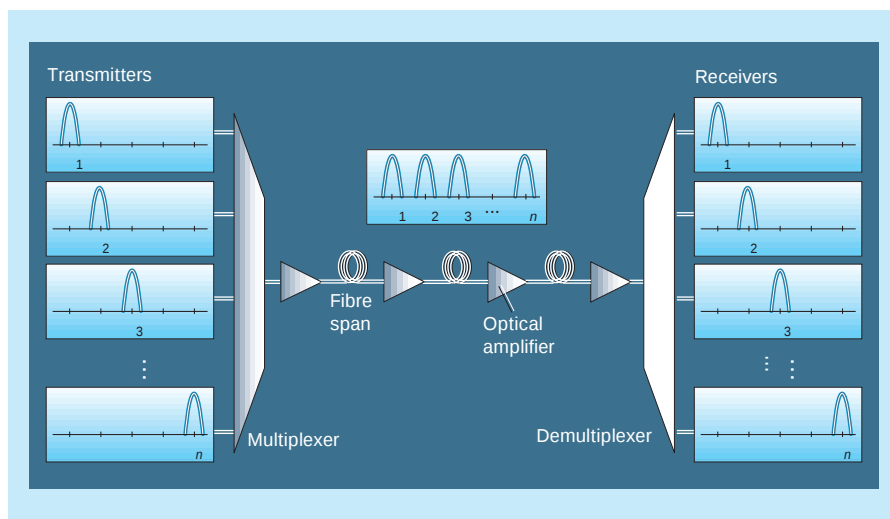


Figure 1 Dense wavelength-division-multiplexed (DWDM) optical-fibre system. Optical signals can be transmitted over very long distances when lengths of fibre (fibre spans) are interspersed with optical amplifiers to boost the signal. The most common amplifiers are made of erbium-doped silica fibres, and provide gain for wavelengths between 1.53 and 1.60 μm . Newer technologies can provide amplification over the entire fibre bandwidth between 1.2 and 1.6 μm . Numerous optical signals at different wavelengths can be combined (by passing them through a multiplexer), transmitted along a common fibre, and separated at the receiving end (using a demultiplexer). The multiplexer and demultiplexer are special types of optical filters. Each signal conveys data at a bit rate R (in bit s^{-1}) and the spacing between adjacent signal wavelengths is $\delta\nu$ (in Hz). The spectral efficiency $R/\delta\nu$ (in $\text{bit s}^{-1} \text{Hz}^{-1}$) measures how efficiently bandwidth is used. Calculations by Mitra and Stark¹ show that fibre spectral efficiency is limited by the noise in the optical amplifiers, and ultimately by the nonlinear properties of the optical fibre.

limited by optical-amplifier noise and by the nonlinear response of the fibre. The effect of the nonlinear response of optical fibres on information capacity has been difficult to evaluate, but Mitra and Stark have succeeded in estimating this fundamental limit.

The classic theory of information developed by Claude Shannon sets fundamental limits on the efficiency of communication². According to Shannon, the capacity of a communications medium is the maximum bit rate that can be transmitted without error, taking into account noise, available bandwidth and limited power. When light is amplified, noise is inherently added, following the principles of quantum mechanics. In addition to amplifier noise, fibre nonlinearities contribute to signal distortion and noise in DWDM systems. Conventional optical fibres have a higher refractive index in their core than in their outer shell, allowing them to guide light by total internal reflection. Fibre materials exhibit a nonlinear response to strong electric fields, such as those of optical signals confined within the small core. Optical amplifiers maintain these fields at high intensities over long distances, and the nonlinear response of the fibre leads to signal distortion and noise.

Mitra and Stark argue that the information capacity of optical fibres in DWDM systems is limited most fundamentally by a nonlinear effect called cross-phase modulation (CPM), in which each signal in the fibre perturbs neighbouring signals. In

current DWDM systems, optical waves are intensity modulated, which means that the intensity (instantaneous power) is varied to encode information — in much the same way that radio waves are amplitude modulated (AM) or frequency modulated (FM). In DWDM systems the full optical bandwidth is occupied by numerous signals at different wavelengths, each of which is modulated separately. The degradation associated with CPM arises because the modulated intensity of each signal perturbs the refractive index of the fibre, thereby distorting the other signals.

Mitra and Stark show that the impact of CPM on the capacity of optical fibres is strongly dependent on a nonlinear intensity scale, I_0 , which is fixed for a particular DWDM system design. When the power transmitted at each signal wavelength is well below I_0 , increasing the power increases the capacity. As the transmitted power approaches I_0 , however, CPM effects increase rapidly, causing the capacity to drop precipitously. So for a given I_0 there is a transmitted power for which the capacity is maximized. The nonlinear intensity scale, I_0 , and the maximum capacity increase with the bandwidth of each signal and the spacing between adjacent signals, and decrease with the total number of signals and the total number of amplifiers. Mitra and Stark express the capacity of a DWDM system in terms of 'spectral efficiency', which is the bit rate per signal divided by the channel



100 YEARS AGO

● I have just tested the inherited powers of swimming in newly hatched pheasants. I find that when placed in tepid water, at the age of about thirty hours, they swim easily with well-coordinated leg-movements and show very little signs of distress.

● Among the many useful publications issued annually by the Royal Observatory of Belgium we draw special attention to "*Ephémérides météorologiques et naturelles*"... The absolute maximum shade temperature was 95.4 °F and the minimum 4.4 °F. The average annual rainfall amounts to 28.56 inches and the mean number of rainy days is 190. The relative humidity at noon is fairly uniform.

● "*Sunny Days at Hastings and St. Leonards*" is the title of a well-illustrated and well-printed little handbook for south-east Sussex, by Messrs. W. H. Sanders and P. Row. There is a 'six-inch' map of Hastings and St Leonards, and another map, on the scale of an inch to four miles, of the country as far as Seaford, Tunbridge Wells and Ashford — all for the price of 6d.
From *Nature* 27 June 1901.

50 YEARS AGO

Polypeptide chains in certain synthetic polymers, in fibrous proteins of the keratin–myosin–fibrinogen group, and also in haemoglobin, appear to be coiled or folded to about half the length of a fully stretched chain. Many different chain configurations have been proposed to account for the X-ray diffraction data, the latest being those of Pauling, Corey and Branson. Until now, however, the lack of any simple and decisive criterion in the X-ray diffraction pattern has made it difficult to test the validity of proposed models. This communication describes a new reflexion, not hitherto observed, which is given by the proteins mentioned above. The spacing at which this reflexion appears excludes all models except the 3.7 residue helix of Pauling, Corey and Branson, with which it is in perfect concord.

This model has two types of repeat: (a) the distance between successive turns of the spiral (5.55 Å), and (b) the spacing along the chain of successive amino-acid residues (1.5 Å). Thus there are in this model 3.7 residues per turn... I have found a new reflexion from planes perpendicular to the fibre axis at a spacing of 1.50 Å, which corresponds to the repeat of the amino-acid residues along the chain.

M. F. Perutz
From *Nature* 30 June 1951.

spacing, and has units of bits per second per hertz ($\text{bit s}^{-1} \text{Hz}^{-1}$).

Reaching the spectral-efficiency limits calculated by Mitra and Stark requires coherent optical detection, which involves interfering the received signal with a reference signal at an identical or slightly different frequency. Current systems use a simpler method, which detects the intensity of the received optical signal. In the absence of amplifier noise and nonlinearities, capacities of optical fibres can be calculated rigorously for systems using intensity detection^{3,4}. We don't know what the limits are for a nonlinear amplified DWDM system using intensity detection, but in the linear regime the maximum spectral efficiency of a DWDM system using intensity detection⁵ is roughly $1 \text{ bit s}^{-1} \text{Hz}^{-1}$ less than half that with coherent detection (for example, 2 and $1 \text{ bit s}^{-1} \text{Hz}^{-1}$ compared with 6 and $4 \text{ bit s}^{-1} \text{Hz}^{-1}$, respectively).

Mitra and Stark¹ now show that for a typical nonlinear DWDM system, assuming coherent detection, the spectral efficiency limit exceeds $3 \text{ bit s}^{-1} \text{Hz}^{-1}$. For comparison, the next-generation commercial DWDM systems are expected to use 40 Gbit s^{-1} channels with 100-GHz spacing — equal to a spectral efficiency of only $0.4 \text{ bit s}^{-1} \text{Hz}^{-1}$. All current DWDM systems use 'on-off keying', a way of encoding bits by the presence or absence of light. Using a binary modulation technique such as on-off keying, spectral efficiency cannot exceed $1 \text{ bit s}^{-1} \text{Hz}^{-1}$ (this limit is reduced to $0.67 \text{ bit s}^{-1} \text{Hz}^{-1}$ by assuming more realistic system parameters). So a system approaching Mitra and Stark's spectral-efficiency limits would require a non-binary encoding technique, such as multilevel intensity or phase modulation.

When calculating the effects of CPM, Mitra and Stark have implicitly assumed a modulation technique involving time-

varying intensity. It is already known that using a constant-intensity modulation technique, such as phase or frequency modulation, can eliminate CPM. The fibre's refractive index varies slightly with the wavelength of light, which tends to convert phase or frequency modulation to intensity modulation. This wavelength dispersion must be carefully compensated for to maintain constant intensity. If we follow the same calculations as Mitra and Stark, and constant intensity is maintained, then the spectral-efficiency limit should increase with transmitted power, in contrast with their results. In reality, as the power increases, spectral efficiency would eventually be limited by other nonlinear effects, such as four-wave mixing.

As Mitra and Stark point out, nonlinearities such as CPM can be cancelled out, in principle, by using a number of clever tricks^{6,7}. Better optical fibres can also help — for example, hollow fibres with air cores have a reduced nonlinear response⁸. In ordinary optical fibres, light can propagate in two orthogonal polarizations — that is, with electric field lines along two perpendicular directions. A simple way to double spectral efficiency is to send two independent

signals with these different polarizations and use polarization-resolved detection. Even without polarization-resolved detection, sending neighbouring signals with perpendicular polarizations is a well-known method to reduce nonlinearities.

Mitra and Stark's work is a useful step towards working out the limits to the spectral efficiency of optical fibres. Nonetheless, our understanding of the ultimate limits, and how to approach them in practice, will continue to evolve.

Joseph M. Kahn is in the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, California 94720, USA.

Keang-Po Ho is in the Department of Information Engineering, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong.

e-mails: jmk@eecs.berkeley.edu

kpho@ie.cuhk.edu.hk

1. Mitra, P. P. & Stark, J. B. *Nature* **411**, 1027–1030 (2001).
2. Shannon, C. E. *Bell Syst. Tech. J.* **27**, 379–423, 623–656 (1948).
3. Kabanov, Y. M. *Theory Prob. Appl.* **23**, 143–147 (1978).
4. Shamai, S. & Lapidot, A. *IEEE Trans. Info. Theory* **39**, 19–29 (1993).
5. Shtaf, M. & Mecozzi, A. in *Proc. Opt. Fiber Commun. Conf.* paper MM1 (Opt. Soc. Am., Washington DC, 2001).
6. Pepper, D. M. & Yariv, A. *Opt. Lett.* **5**, 59–60 (1980).
7. Pare, C. *et al. Opt. Lett.* **21**, 459–461 (1996).
8. Cregan, R. F. *et al. Science* **285**, 1537–1539 (1999).

Immunology

T cells and tumours

Drew Pardoll

The immune system's response to new tumours is complicated but seems to depend on where and when tumours develop. This will have to be much better understood to enlist a patient's immune defences in fighting cancer.

Does the immune system see tumours more as 'self' or as 'foreign'? The answer is complex, as is plain from papers published here in April¹ and on page

1058 of this issue². They show that the immune response probably depends on the location and properties of tumour cells early in their development.

Quantum engineering

Catch the atom

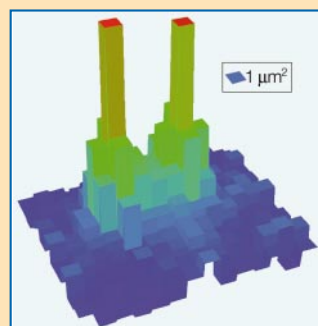
Two groups have independently succeeded in confining single atoms in microscopic traps (N. Schlosser *et al.*, *Nature* **411**, 1024–1027; 2001, and S. Kuhr *et al.*, *Science*, 14 June 2001, 10.1126/science.1062725). This feat opens the way to designing experiments in which various quantum-mechanical effects can be exploited. Several techniques already exist to manipulate individual particles, such as photons and ions, but it has been notoriously difficult to pin down neutral atoms.

The two groups developed methods to attract cooled-down atoms towards spots of high electric-field intensity in traps formed by laser beams. Schlosser *et al.* find that the individual rubidium atoms entering their traps scatter enough photons to image them with a CCD camera. The images show that the traps are occupied by either no atoms or just one at a time. Moreover, once an atom is caught, it can be kept trapped for up to 2 seconds — a veritable lifetime in quantum mechanics.

Kuhr *et al.* exerted their control over chilled caesium atoms and designed traps that can catch any desired small number of atoms. Intriguingly, they also found that these atoms can be catapulted into free flight from the trap, raising prospects of an 'atom on demand' delivery service.

As a final flourish, Schlosser *et al.* have positioned two traps, each containing a single atom, close to each other as shown here in the CCD image. In this set-up, the internal states of the two atoms can be intimately related

to each other, or 'entangled'. Such entangled states can be used as logic elements for efficient computation tasks. Liesbeth Venema



First, some background. In the 1950s, the immune-surveillance hypothesis was formulated. According to this idea, the immune system surveys the body for specific antigens expressed by newly arising tumours, most of which are rapidly eliminated. Progressive cancer was seen as a rare occurrence in which tumours evade immune surveillance, ostensibly through specific resistance or 'cloaking' mechanisms. This view was supported by findings that cancers have various ways of making themselves resistant to the immune system, one tactic being inactivation of an afflicted cell's major histocompatibility complex (MHC)^{3,4}, or its associated machinery^{5,6}. This would allow the cell to evade recognition by T cells. Recognition of diseased cells and their execution by T cells, after presentation by the MHC of foreign antigens at the cell surface, is one of the main immune defence mechanisms.

The immune-surveillance hypothesis ran into trouble, however, as various findings emerged. The common, non-virus-associated cancers (breast, lung and so on) do not arise any more frequently in people with a defective immune system than in those with a fully functioning system⁷. Likewise, nude mice — which have impaired T-cell and B-cell (antibody) immunity — have tumour incidences equivalent to those of normal mice⁸. Finally, studies of transgenic mice, in which specific T cells for tumour-associated antigens could be marked and monitored, showed that tumour cells can induce immune tolerance^{9–11}. In some cases, T cells were activated, but only transiently. In others, the tumour-specific T cells were rendered unresponsive (anergic). These findings all support an opposite view to that of the immune-surveillance hypothesis — that the immune system sees tumours more as self than as foreign.

The new papers by Shankaran *et al.*¹ and Ochsenbein *et al.*² show that the truth lies somewhere between these two views. Shankaran *et al.* evaluated the incidence of tumour formation in immunodeficient mice that provide 'cleaner' subjects than the nude strain (which have residual T- and B-cell function). The authors used mice in which the RAG and/or STAT-1 genes had been deleted, enabling them to analyse responses of the innate arm of the immune system as well as the adaptive (antigen-mediated) arm that depends on T and B cells. RAG 'knockout' mice cannot rearrange the genes that are central to adaptive immunity. STAT-1 knockout mice lack the pathways, mediated by γ -interferon, that are essential for both the adaptive and innate responses.

Cancer incidence in both of these types of mice was clearly higher than in normal mice. But the type and location of tumours depended on which components of the

immune response were absent. RAG-deficient mice largely suffered from cancer of epithelial tissue in the intestine. Double RAG/STAT-1 knockouts developed analogous breast cancers as well, so innate immunity may suppress tumours at certain sites. Both types of mice particularly developed spontaneous tumours from around one year of age and beyond (this, by mouse standards, is senior-citizen status). So, even without important elements of the immune

system, the mice did pretty well, showing that its role in limiting tumorigenesis is more subtle than implied by the classic surveillance hypothesis.

Shankaran and colleagues' findings support a highly modified version of immune surveillance in which the immune system can act as an 'extrinsic suppressor' of tumours in certain locations or tissues. This is analogous to the classic intrinsic suppressors, such as that encoded by the *p53* gene,

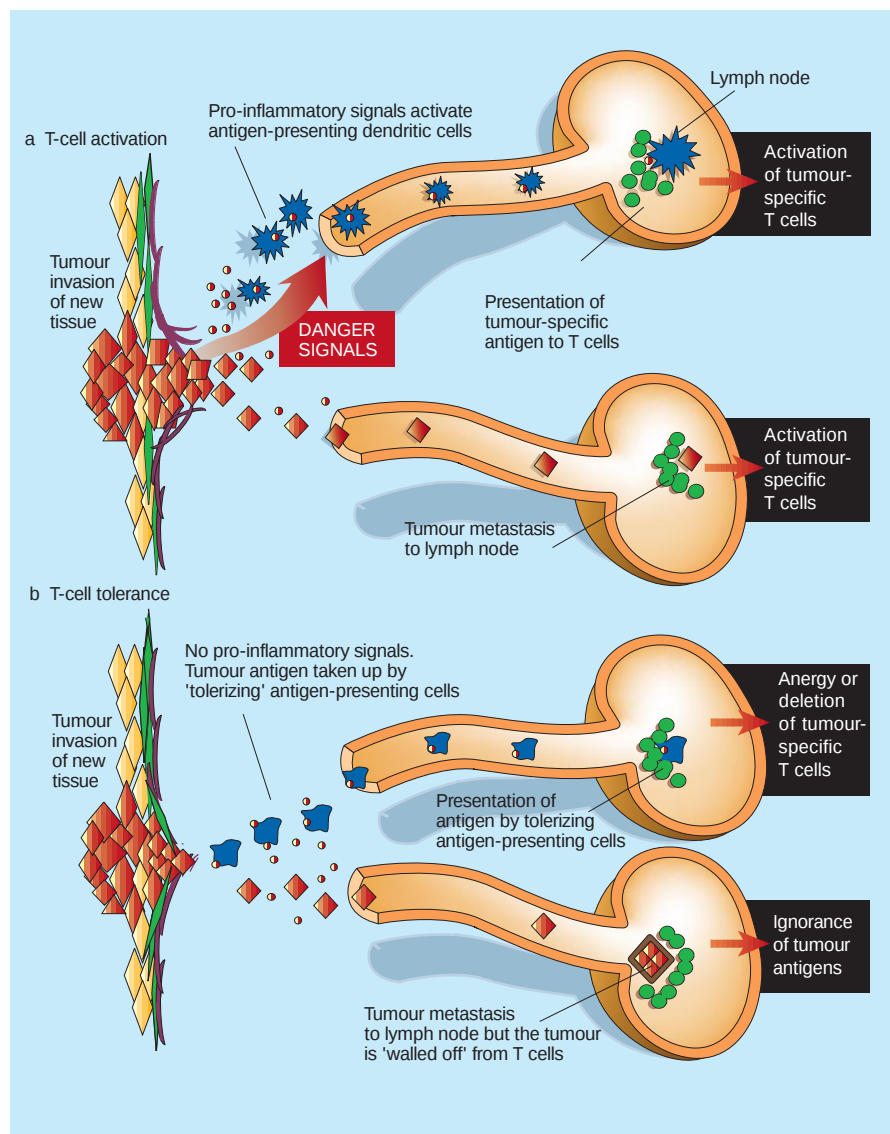


Figure 1 T-cell responses to tumours in different circumstances. Tumour invasion across normal tissue barriers and metastasis can lead to a, T-cell activation, or b, T-cell tolerance. a, As shown in the upper pathway, the tumour may disrupt the intercellular matrix, which can result in the generation of pro-inflammatory signals. These signals cause progenitor cells from the bone marrow to differentiate into antigen-presenting dendritic cells. When carried into the lymph nodes, the dendritic cells present the tumour-specific antigen to T cells, which become activated. Alternatively (a, lower pathway), some tumours may metastasize to lymph nodes and directly activate T cells. b, Some tumours can invade new tissue and metastasize without generating pro-inflammatory signals. Here (upper pathway), 'tolerizing' antigen-presenting cells can take up tumour antigen and, after passing into lymph nodes, present the tumour-specific antigen to T cells and make them tumour-tolerant in one of two ways: by anergy (in which the T cells are rendered unresponsive) or deletion (in which the T cells are destroyed). Alternatively (b, lower pathway), although a tumour may metastasize to the lymph nodes, it becomes 'walled off' from T cells, which therefore remain ignorant of the tumour antigens.

which must be inactivated or bypassed to allow tumour progression.

In their paper, Ochsenbein *et al.*² describe how they have correlated the ability of tumours to migrate into secondary lymphoid organs — lymph nodes and spleen — with the activation of T-cell responses against tumour antigens. The significance of these organs is that they are the main places where T cells await activation.

Ochsenbein *et al.* looked at several different mouse tumours and found that some metastasizing tumours that fail to 'seed' lymph nodes and spleen are 'ignored' by the immune system. They also found that when tumours were injected into secondary lymphoid organs, they segregated into two types. One type — 'immunogenic' tumours that elicited a strong immune response — circulated in lymphoid regions, intermingling with T cells and allowing direct antigen presentation to T cells and activation of anti-tumour immunity. Another, indirect pathway to T-cell activation is well characterized (Fig. 1a, upper pathway): antigen-presenting cells known as dendritic cells take up tumour antigens, carry them to lymph nodes and present them to T cells^{12,13}. The circumstances under which the direct or indirect pathway predominates remain to be identified.

The second type of tumour identified by Ochsenbein *et al.* was only weakly immunogenic. This type grew as nodules, 'walled off' from the immune system by barriers that prevented activation of anti-tumour immunity. So even if tumour antigens make their way to secondary lymphoid organs, this is not necessarily enough to activate immunity. It is indeed likely that this second kind of tumour represents most human tumours. As any oncologist will tell you, lymph-node metastases are usually a poor prognostic sign, indicating a high likelihood of cancer spread to elsewhere. Presumably, tumours that directly activate T cells in lymph nodes are eliminated before becoming clinically detectable.

How, then, can one reconcile the evidence that tumours can either activate the immune system or make it tolerate them? Possible answers are outlined in Fig. 1. At certain points in their development, tumours must become invasive and create their own blood supply. These severe disruptions of cellular processes and tissue architecture may provoke pro-inflammatory signals, which can convert immune tolerance to activation through induction of dendritic cells¹⁴ (Fig. 1a, upper pathway). Dendritic cells can ingest tumour antigens; after migration to the lymph nodes, and with certain co-stimulatory molecules, they can activate T cells quite efficiently. Alternatively, tumour cells may migrate to lymph nodes and activate T cells directly (Fig. 1a, lower pathway).

In other circumstances, despite causing serious disruption, it seems that a tumour may fail to generate pro-inflammatory signals. The immune system then probably views a tumour as 'self' tissue and tolerates its antigens. This tolerance can take the form of 'ignorance', as outlined by Ochsenbein *et al.*², in which the immune system is never made aware of the tumour (Fig. 1b, lower pathway). Or it can result from active processes such as T-cell anergy or deletion (Fig. 1b, upper pathway). These processes involve transfer of tumour antigens to 'tolerizing' antigen-presenting cells, which develop in bone marrow, carry antigens to lymph nodes and, in the absence of co-stimulatory signals, present them to T cells. The result is T-cell anergy or deletion^{15–17}.

Against this background, the ultimate immune response may depend on where and when tumour-specific antigens form. If the tumour does not generate a new antigen during the pro-inflammatory phases of its development, then immune tolerance dominates. But if the tumour is unlucky enough (from the tumour's perspective) to generate a strong antigen just when it is generating a pro-inflammatory response, the outcome will be a vigorous anti-tumour response. These tumours will be eliminated naturally unless they have developed specific resistance mechanisms, such as downregulation of MHC or the antigen-processing machinery.

Overall, we can conclude that tumours that reach the stage of being clinically detectable are likely to have done so in one of two ways. They will either have generated tolerance in the immune system or have developed ways of resisting immune recognition. In terms of cancer treatment, we need to identify ways to break tolerance or circumvent resistance mechanisms. ■

Drew Pardoll is in the Department of Oncology, Medicine, Pathology, Molecular Biology and Genetics, Johns Hopkins University, Baltimore, Maryland 21231, USA.

e-mail: dmpardol@jhmi.edu

1. Shankaran, V. *et al.* *Nature* **410**, 1107–1111 (2001).
2. Ochsenbein, A. F. *et al.* *Nature* **411**, 1058–1064 (2001).
3. Travers, P. J., Arklie, J. L., Trowsdale, J., Patillo, R. A. & Bodmer, W. F. *Natl Cancer Inst. Monogr.* **60**, 175–180 (1982).
4. Marincola, F. M., Jaffee, E. M., Hicklin, D. J. & Ferrone, S. *Adv. Immunol.* **74**, 181–273 (2000).
5. Trowsdale, J., Travers, P., Bodmer, W. F. & Patillo, R. A. *J. Exp. Med.* **152**, 11–17 (1980).
6. Restifo, N. P. *et al.* *J. Exp. Med.* **177**, 265–272 (1993).
7. Groopman, J. E. *Semin. Oncol.* **14**, 1–6 (1987).
8. Stutman, O. *J. Natl Cancer Inst.* **2**, 353–358 (1979).
9. Lauritzen, G. F., Weiss, S., Dembic, Z. & Bogen, B. *Proc. Natl Acad. Sci. USA* **91**, 5700–5704 (1994).
10. Speiser, D. E. *et al.* *J. Exp. Med.* **186**, 645–653 (1997).
11. Staveley-O'Carroll, K. *et al.* *Proc. Natl Acad. Sci. USA* **95**, 1178–1183 (1998).
12. Huang, A. Y. *et al.* *Science* **264**, 961–965 (1994).
13. Toes, R. E. *et al.* *Cancer Res.* **56**, 3782–3787 (1996).
14. Fuchs, E. J. & Matzinger, P. *Semin. Immunol.* **8**, 271–280 (1996).
15. Adler, A. J. *et al.* *J. Exp. Med.* **187**, 1555–1564 (1998).
16. Miller, J. F. *et al.* *Immunol. Rev.* **165**, 267–277 (1998).
17. Sotomayor, E. *et al.* *Blood* (in the press).

Daedalus

Phones and the brain

One of the problems of using a mobile phone is that the brain is sensitive to microwaves. Daedalus points out that each brain cell has many inputs (dendrites) and one main output (its axon). All these fibres go to other brain cells. Nerve pulses from a specific combination of dendrites fire (or inhibit the firing of) an impulse down the axon. So each brain cell acts as a logical gate; it learns from experience. It alerts the dendrites which receive each pulse, so that they fire or inhibit its resulting output. This changing sensitivity stores our memories and abilities.

It is reasonable that the substances which do all this are proteins. They seem well adapted as the ultimate storers of knowledge. A protein molecule may, in one configuration, inhibit a brain cell; in another, it may potentiate it. And the energies that separate the protein configurations are in the microwave region, below the infrared band which holds bond-stretching itself.

This theory, says Daedalus, explains the great plasticity and learning capacity of the brain, and also its sensitivity to microwaves. Reconfiguration of protein structure must require the absorption of microwaves to shift the amino-acid moieties from one configuration to another. Hence the fears that children educated under microwave towers, and mobile-phone users, may suffer memory loss. A single frequency, as from a mobile phone, could at best scramble only one or two protein transitions; it could still leave gaps in a memory. The more complex transmissions from a microwave tower, at many frequencies, could be more troublesome.

So Daedalus is exploiting the recent culturing of nerve cells in a special medium. His idea is that in a Petri dish the cells can be exposed to much stronger microwaves than in any brain. The way they link together by their dendrites, or draw apart, will reveal the effects of the microwaves. Some frequencies may turn out to be particularly dangerous. They may trigger many changes of state in the dendrites. Other frequencies may turn out to be well clear of the brain's operating frequencies. These will be the ones to use for mobile-phone use. And, of course, a simple design change will reduce the phone's impact on its user by a factor of about four. Put the irradiating antenna at the microphone end, projecting from the jaw-line, well clear of the ear and the brain.

David Jones

Night-time predation by Steller sea lions

New insight into the feeding habits of these mammals will help conservation attempts.

Measures have been taken to curtail commercial fishing of walleye pollock (*Theragra chalcogramma*) in Alaska in an attempt to stop the decline of its endangered population of Steller sea lions (*Eumetopias jubatus*). But our night-time observations of these mammals in Prince William Sound using infrared scanning technology, combined with acoustic surveillance of their prey's behaviour, reveal that the sea lions feed exclusively on Pacific herring (*Clupea pallasii*), which are less abundant than pollock but are found closer to the surface at night.

Food limitation is the principal factor in the decline of Steller sea lion populations^{1–4}. This decline could be explained by competition with commercial fisheries, as it has coincided with the growth of the pollock-fishing industry, which has become one of the largest fisheries in the world, or it could be related to a change in predator–prey relationships, possibly driven by ocean climate shifts⁵. Central to the uncertainty surrounding the drop in the numbers of Steller sea lions is a lack of observational data on their foraging ecology. There is no quantitative information available that directly relates the foraging behaviour of these animals to the abundance of prey species.

During the winter period, nutritional stress is high. Sonar surveys^{6,7} of the abundance and distribution of adult Pacific herring and walleye pollock in winter have been made in Prince William Sound in Alaska since the early 1990s⁸. Steller sea lions were seen during the day near herring schools, but as no foraging activity was detectable, the significance of this co-occurrence was questionable.

We complemented our sonar surveys during March 2000 with infrared scanning of the Steller sea lions. This technology, which is widely used in night-time military operations and surveillance, enabled us to monitor the animals' activity during the hours of darkness. Our system had a 27° × 18° field of view and a spectral response of 7–14 µm.

The estimated herring biomass in Prince William Sound in the sonar survey of March 2000 was 7,281 metric tonnes (95% confidence interval, 5,898–8,664). The estimate of pollock biomass was 28,277 metric tons (95% confidence interval, 26,034–30,420). Despite the much greater abundance of pollock, the infrared system revealed that foraging by Steller sea lions was exclusively on herring and was conducted only at night. Foraging activity was intense on dense herring schools (Fig. 1).

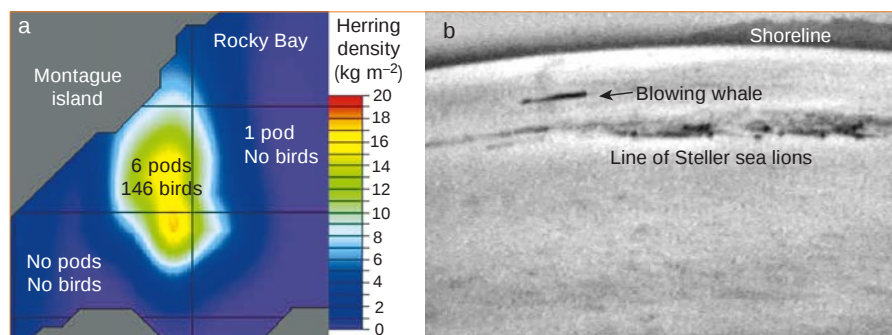


Figure 1 Location of groups (pods) of Steller sea lions around herring schools. **a**, Combined acoustic and infrared sensors reveal sea lions and birds located on the surface above the herring school at night in Rocky Bay, Prince William Sound (March 2000). **b**, Infrared video image showing a line of Steller sea lions and a humpback whale on the sea surface above a school of herring.

Steller sea lions were often observed swimming side by side in a row of 50 or more individuals along the edges of a school, suggesting that they were herding the herring. Humpback whales and seabirds were also seen to be feeding alongside the sea lions (Fig. 1). By contrast, no sea lions were coincident with pollock schools.

The sonar records revealed herring schools at depths of 10–35 m at night, but deeper during the day. Walleye pollock, on the other hand, remained at depths of over 100 m during both day and night. Pollock schools were also found in less protected regions and were further offshore. Although Steller sea lions are capable of dives exceeding 250 m (ref. 9), the more accessible distribution of herring at night may be the primary factor in the foraging behaviour of the sea lions. This distribution of herring is characteristic during an extended overwintering period in the North Gulf of Alaska.

Our results indicate that the dependence of Steller sea lions on herring as prey has been underestimated. The infrared scanning technology that has led us to this con-

clusion should also help in the evaluation of night-time foraging behaviour of other marine mammals and seabirds, with its remarkable ability to detect individual fish flipping on the sea surface at a distance of 5–30 m, as well as sea lions, whales and birds at over 100 m.

Gary L. Thomas, Richard E. Thorne

Prince William Sound Science Center, PO Box 705, Cordova, Alaska 99574, USA
e-mail: thorne@pwssc.gen.ak.us

1. National Research Council *The Bering Sea Ecosystem* (National Academic, Washington DC, 1996).
2. Castellini, M. Alaska Sea Grant College Program AK-SG-93-01, 4–13 (1993).
3. Merrick, R. L. et al. *Can. J. Fish Aquat. Sci.* **34**, 1342–1348 (1997).
4. Loughlin, T. R. *Biosphere Conserv.* **1**, 91–98 (1998).
5. Mantua, N. J. et al. *Bull. Am. Meteorol. Soc.* **78**, 1069–1079 (1997).
6. MacLennan, D. N. & Simmonds, E. J. *Fisheries Acoustics* (Chapman & Hall, London, 1992).
7. Thorne, R. E. et al. *FAO Fish. Rep.* **300**, 217–222 (1983).
8. Thomas, G. L. et al. in *Developing and Sustaining World Fisheries Resources — the State of Science and Management* (eds Hancock, D. A. et al.) 606–613 (CSIRO, Collingwood, Australia, 1997).
9. Merrick, R. L. et al. *Polar Res.* **13**, 105–114 (1994).

Genomics

Genes lost during evolution

One of the main conclusions presented by the International Human Genome Sequencing Consortium is that “hundreds of genes appear to have resulted from horizontal gene transfer from bacteria at some point in the vertebrate lineage”¹. We noticed that a significant proportion of these human genes have closely related orthologues in the primitive eukaryote *Dictyostelium*. This observation supports independent gene loss in multiple lineages (worm, fly, yeast, plants) rather than hori-

zontal gene transfer from bacteria.

The human genome sequence revealed 113 genes that share a high degree of identity with bacterial genes, but are absent in the completely sequenced genomes of *Caenorhabditis elegans*, *Drosophila melanogaster*, *Saccharomyces cerevisiae* and *Arabidopsis thaliana*¹. Do these genes represent examples of horizontal gene transfer from bacteria to the vertebrate lineage, or were they present in both prokaryotes and early eukaryotes, but subsequently lost from all non-vertebrate eukaryotic lineages? Although this latter possibility may seem unlikely, we recently identified a gene in *Dictyostelium* that is clearly an orthologue of the gene that encodes soluble

adenylyl cyclase in bacteria and vertebrates, but has not been identified in other eukaryotes². *Dictyostelium* is located in the evolutionary tree between plants and the fungi/animal crown³, and sequencing of its genome is approaching completion⁴ (see also <http://dictybase.org>).

We used all 113 listed human genes to screen for homologous sequences in *Dictyostelium* (27 February 2001; see supplementary information). A TBLASTN screen of the *Dictyostelium* database yielded 36 sequences with expectation values of less than 10^{-10} . BLASTX analysis with the obtained *Dictyostelium* DNA sequences against GenBank identified 11 genes that represent clear *Dictyostelium* orthologues of human genes: the human sequences share a higher degree of identity with *Dictyostelium* than with bacterial sequences, and the bacterial sequences score more highly with respect to *Dictyostelium* than they do to humans (on the basis of BLAST expectation values). A further 17 *Dictyostelium* sequences share a high degree of identity with the human sequence, but are not obvious intermediates between the bacterial and vertebrate orthologues (see supplementary information). Thus, in at least 11 cases, the *Dictyostelium* and human genes have a common ancestor, eliminating the need to invoke horizontal gene transfer from bacteria.

One of the human proteins with an orthologue in *Dictyostelium* is monoamine oxidase (MAO). Phylogenetic analysis of this enzyme reveals a gene duplication late in the vertebrate lineage (MAO-A and MAO-B in Fig. 1). These paralogs seem to share a predecessor with *Dictyostelium*, indicating that monoamine oxidase was present in early eukaryotes, and implying that the gene has been lost in worm, fly,

plants and yeast.

Within the group of 113 genes proposed to have entered the human genome by horizontal gene transfer from bacteria, we have identified at least 11 that probably arose through normal evolution with gene loss in several lineages, suggesting that gene loss is not a rare event. With several ongoing genomic sequencing projects for lower eukaryotes, it will be interesting to see how many genes have truly undergone horizontal transfer.

Jeroen Roelofs, Peter J. M. Van Haastert

Department of Biochemistry, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

e-mail: p.j.m.van.haastert@chem.rug.nl

1. International Human Genome Sequencing Consortium. *Nature* **409**, 860–921 (2001).

2. Roelofs, J. et al. *Biochem. J.* **354**, 697–706 (2001).

3. Baldauf, S. L. et al. *Science* **290**, 972–977 (2000).

4. Kay, R. R. & Williams, J. G. *Trends Genet.* **15**, 294–297 (1999).

Supplementary information is available at <http://www.nature.com> or as paper copy from the London editorial office of *Nature*.

Bone-marrow transplantation

Failure to correct murine muscular dystrophy

Bone-marrow cells have the potential to differentiate into other cell types such as muscle fibres, and can be transplanted into acutely¹ or chronically² damaged muscle as a way of delivering normal dystrophin (the protein that is defective or missing in Duchenne's muscular dystrophy) to the skeletal and heart muscle of *mdx* mice^{2,3}, an animal model for this disease. But the corrective potential of this approach has been hard to estimate against the high background of muscle fibres that spontaneously revert to synthesizing dystrophin, a feature of the original *mdx* mutation⁴. Here we test the long-term efficacy of bone-marrow transplantation in a different *mdx* mutant which is free of this problem and find that it has no impact on murine muscular dystrophy.

The *mdx4cv* mutant (in which a C-to-T nucleotide transition generates a stop codon in exon 53 of the dystrophin gene) has almost no background of revertant fibres in skeletal muscle⁴. We sublethally irradiated (900 cGy) a group of 15 8-week-old *mdx4cv* mice (C57Bl/6/Ly-5.2 background) and transplanted them with a total of 1.5×10^7 bone-marrow cells from a pool of 6-week-old, co-isogenic (C57Bl/6/Ly-5.1) animals. We killed the mice at regular intervals from 9 weeks to 10 months after transplantation, and monitored the engraftment of donor cells by cytofluorimetric analysis of the proportion of Ly-5.1 marker compared with Ly-5.2. The degree of chimaerism averaged $85 \pm 2.7\%$ in bone marrow (mean $\pm$ s.e.m.), $93 \pm 1.1\%$ in

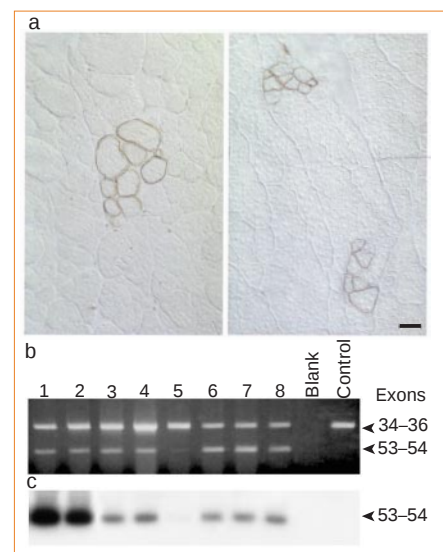


Figure 1 Expression of dystrophin in *mdx* mice 6 months after transplantation with bone-marrow cells from co-isogenic, normal donors. **a**, Immunohistochemical staining of frozen sections of tibialis muscle with an anti-dystrophin monoclonal antibody. Scale bar, 50 μ m. **b**, Detection of wild-type dystrophin mRNA by RT-PCR amplification using specific primers for exons 53 and 54 in samples of total RNA extracted from skeletal muscle of transplanted *mdx* mice (lower bands). A fragment encompassing exons 33–36 in both wild-type and mutant dystrophin RNA is amplified as an internal control (upper bands). Lanes 1–8, samples from transplanted mice; 'blank', PCR assay without RNA; 'control', mock-transplanted *mdx* control. **c**, Southern-blot hybridization with an internal, specific oligonucleotide probe for exon 53.

spleen, $92 \pm 2.9\%$ in thymus and $94 \pm 0.8\%$ in peripheral blood throughout the follow-up study.

We counted dystrophin-positive (dys^+) fibres in histological sections of representative muscles (tibialis anterior, quadriceps, diaphragm) after immunohistochemical staining with an anti-dystrophin antibody in transplanted and age-matched, mock-transplanted, control *mdx4cv* mice. Clusters of dys^+ fibres were apparent in muscle sections of transplanted animals, averaging $0.23 \pm 0.05\%$ (minimum, 0.06%; maximum, 0.54%) throughout the 10-month study (Fig. 1a). The proportion of dys^+ fibres in control animals averaged $0.14 \pm 0.03\%$ (minimum, 0.02%; maximum, 0.33%), a statistically significant difference ($F = 5.99$, $P = 0.02$). In neither group was there any significant increase in the number of dys^+ fibres in young (under 5 months) and old (over 12 months) animals. The average number of fibres contained in each dys^+ cluster varied from 3 to 30, with no significant change with age in either group.

To demonstrate the presence of normal dystrophin in the muscle of transplanted mice (the antibody does not distinguish between corrected and revertant fibres), we developed a polymerase chain reaction with reverse transcription (RT-PCR) assay to distinguish wild-type dystrophin messen-

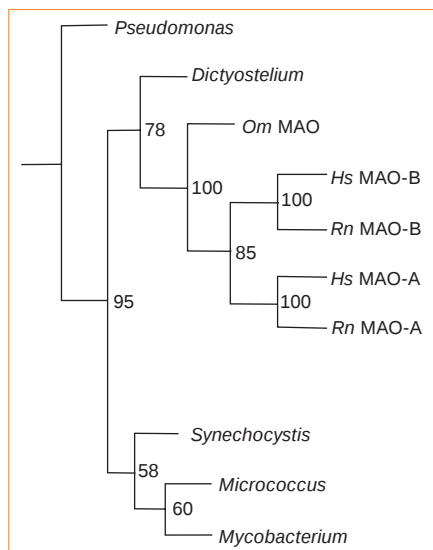


Figure 1 Phylogenetic analysis of monoamine oxidase (MAO). Numbers indicate values of bootstrap analysis ($n = 100$). Hs, *Homo sapiens*; Rn, *Rattus norvegicus* (rat); Om, *Oncorhynchus mykiss* (rainbow trout).

ger RNA from that carrying the *mdx4cv* mutation. As shown in Fig. 1b, c, a 5' primer specific to the wild-type sequence allowed amplification of a 240-base-pair fragment encompassing exons 53 and 54 of mRNA extracted from muscle tissue of control C57Bl/6 and all transplanted animals, but not from mRNA from mock-transplanted *mdx4cv* mice. All PCR reactions were internally normalized by simultaneous amplification of a 378-base-pair fragment encompassing exons 33–36.

Our results indicate that repair by bone-marrow-derived cells never exceeds 1% of total muscle fibres during the lifespan of transplanted *mdx* mice. Transplantation of total bone marrow, or even of highly enriched stem-cell fractions from bone marrow², therefore has a very limited impact on muscle-cell replacement, and no ameliorating effect on murine muscular dystrophy. This might be explained by poor recruitment of bone-marrow cells to the dystrophic muscle, or by bone-marrow-derived cells having no selective advantage over resident myogenic precursors, which regenerate muscle throughout life in *mdx* mice.

We do not address strategies for overcoming these limitations here, although they are likely to be the focus of future research. It should be stressed that the *mdx* mouse model does not reproduce the clinical picture of Duchenne's muscular dystrophy in humans, in which muscle degeneration gives rise to fibrosis rather than to repair⁵, so it remains possible that dystrophin-producing cells derived from bone marrow might be subject to positive selection in a human context.

Giuliana Ferrari*, **Anna Stornaiuolo†**,
Fulvio Mavilio*‡

*H. San Raffaele-Telethon Institute for Gene Therapy, and †GenEra SpA, Via Olgettina 58, 20132 Milano, Italy
e-mail: f.mavilio@hsr.it

‡Department of Biomedical Sciences, University of Modena School of Medicine, 41100 Modena, Italy

1. Ferrari, G. *et al.* *Science* **279**, 1528–1530 (1998).
2. Gussoni, E. *et al.* *Nature* **401**, 390–394 (1999).
3. Bittner, R. E. *et al.* *Anat. Embryol.* **199**, 391–396 (1999).
4. Danko, I. *et al.* *Pediatr. Res.* **32**, 128–131 (1992).
5. Emery, A. E. H. *Duchenne Muscular Dystrophy* 2nd edn (Oxford Univ. Press, Oxford, 1993).

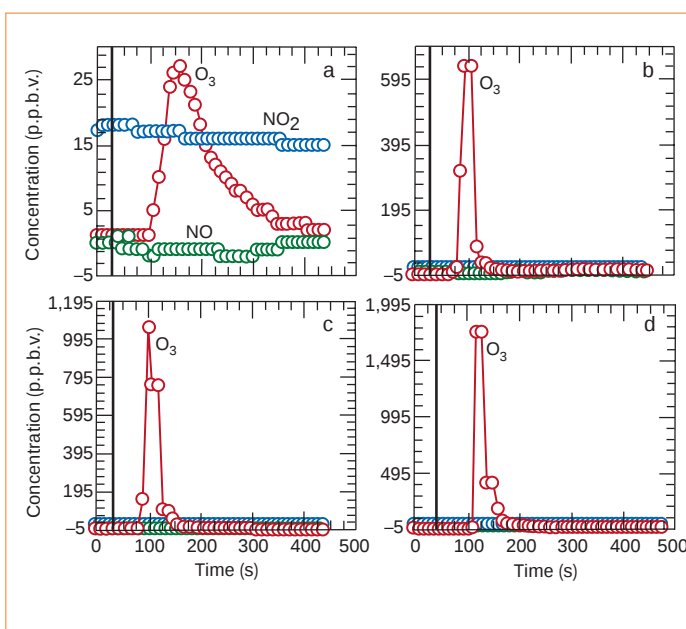
Microclimate

Formation of ozone by fireworks

Ozone is a secondary pollutant and greenhouse gas that is formed from molecular oxygen in the presence of sunlight and nitrogen oxides. The extent of production also depends on the presence of volatile hydrocarbons, carbon monoxide and methane^{1–6}. But we have discovered a surprising source of ozone which is generated in spontaneous bursts even in the

Figure 1 Ambient concentration (parts per billion by volume) of O₃ (red circles), NO (green) and NO₂ (blue) in the environment as a function of time after lighting colour-emitting 'pencil' sparklers. **a–d**, Values correspond to the simultaneous burning of one, three, four and six sparklers, respectively; vertical black lines indicate the duration of burning for the sparklers (starting at 0 s). O₃ begins to accumulate in each case at around 100 s; this time lag corresponds to the volume of air in the assembly of the Teflon tube and manifold attached to the O₃ and NO_x analysers. During the experiments, the open end of the

Teflon tube leading to the analysers was kept 900 cm away from the sparklers. The average weight of inflammable material (initial weight minus weight after burning) was 9.15 ± 0.85 g per sparkler. The burn time varied from 26 s for one sparkler to a maximum of 42 s for six; this variation was due to differences in the weight of the burning mixture. O₃ formation was independent of microclimatic factors such as temperature and humidity of the ambient air. The concentrations of NO and NO₂ did not change during the O₃ burst. Further details are available from the authors.



absence of sunlight and nitrogen oxides — namely, the exuberant mass of colour-emitting sparklers that are lit during the Diwali festivities, which take place every year during October and November in Delhi, India. The underlying process of ozone formation resembles that induced by ultraviolet radiation in the stratosphere^{7,8}.

We undertook a routine, real-time monitoring of the concentrations of NO_x (NO and NO₂), O₃ and other microclimatic factors at a known pollutant-receptor site⁹ in Delhi in order to determine the effects of burning unprecedented numbers of fireworks on the local environment during the festive period in November 1999. We found that during the festival period the ozone concentration peaked at around noon and fell to negligible levels after sunset.

On Diwali night (7 November), a small build-up of O₃ (9 ± 1 parts per billion by volume) was detected between 20:40 and 02:30 hours (results not shown). During this period, no correlation was found between NO_x concentration and O₃ formation, indicating that the ozone was unlikely to have been generated in reactions involving ambient NO_x. This observation was surprisingly different from night-time O₃ measurements obtained on the other dates, when no O₃ formation was detected.

Further experiments carried out under different climatic conditions showed that there is a linear regression between the total amount of inflammable material present in sparklers and the cumulative O₃ formed (correlation, 0.993). There was no change in ambient NO_x concentration before,

during or after these experiments (Fig. 1).

Sparklers depend on a combination of different metal salts to generate their colour and sparkle¹⁰ — these include potassium perchlorate, sulphur, strontium nitrate, barium nitrate, sodium oxalate, calomel, aluminium and manganese. When burnt, a significant proportion of the light emitted by these constituents has a wavelength below 240 nm (ref. 11). The radiative energy of these emissions is sufficient to dissociate atmospheric molecular oxygen into atomic oxygen, enabling the reaction $O_2 + O \rightarrow O_3$ to take place. This proposed mechanism could explain the formation of bursts of O₃ without the participation of NO_x, and is therefore similar to the process of ultraviolet-radiation-induced formation of O₃ in the stratosphere^{7,8}.

Arun K. Attri, **Ujjwal Kumar**, **V. K. Jain**
School of Environmental Sciences, Jawaharlal Nehru University, New Delhi 110 067, India
e-mail: attriak9@bol.net.in

1. Ramanathan, V., Singh, H. B., Ciceron, R. J. & Kiehl, J. T. *J. Geophys. Res.* **90**, 5547–5566 (1985).
2. Lalic, A. A., Weubbles, D. J. & Logan, J. A. *J. Geophys. Res.* **95**, 9971–9981 (1990).
3. Wang, W. C. & Sze, N. D. *Nature* **286**, 589–590 (1980).
4. Crutzen, P. J. & Zimmermann, P. H. *Tellus* **43A–B**, 136–151 (1991).
5. Walcek, C. J. & Yuan, H.-H. *J. Appl. Meteorol.* **34**, 1056–1069 (1995).
6. Johnson, D. W. *et al.* *Nature* **353**, 617–621 (1991).
7. Chapman, S. R. *Meteorol. Soc. Mem.* **3**, 103–125 (1930).
8. Brausser, G. P. *et al.* *J. Geophys. Res.* **95**, 5639–5655 (1990).
9. Singh, A. *et al.* *Atmos. Environ.* **31**, 3421–3427 (1997).
10. Holmes, H. H. in *The Encyclopedia Americana* Int. edn, Vol. 11, 263–265 (American Corp., New York, 1983).
11. Reader, J. & Corliss, C. H. in *CRC Handbook of Chemistry and Physics* 63rd edn (eds Weast, R. C. & Astle, M. J.) E205–E368 (CRC, Florida, 1982–83).

Laminar flows

Subcellular positioning of small molecules

Localized perturbation of processes that take place inside the living cell depends on molecular and spatial discrimination on a micrometre scale. Here we report the use of multiple laminar streams in a microfluidic channel to deliver membrane-permeable molecules to selected subcellular microdomains. This technique opens up avenues for non-invasively visualizing, probing and manipulating the cellular metabolic and structural machinery.

We have applied this method (called 'partial treatment of cells using laminar flows', or PARTCELL) to study the subcellular processes of mitochondrial movement and changes in cytoskeletal structure. Mitochondrial positioning is important for localized production of metabolic energy within different cellular microdomains, but analysis has so far been limited to imaging the movement of individual organelles. Local alterations in cytoskeletal architecture help to control cell shape, growth and motility^{1,2}, but investigation of these has previously had to rely on the exposure of whole cells to drugs that cause gross modification, such as complete disruption of actin microfilaments.

Microfluidic systems were prepared by placing a polydimethylsiloxane slab with channel features (300 × 50 µm) moulded into its surface onto a glass coverslip (Fig. 1a)^{3–5}. By allowing different solutions to flow from the inlets at low velocity (about 0.6 cm s⁻¹), parallel streams of different liquids are created in the main channel (Fig. 1a, b)^{4,6,7}. Under these conditions, there is no turbulence and the streams flow next to each other without mixing (apart from a small amount of diffusion). The width of each stream and the position of the interface between adjacent streams can be controlled by adjusting the relative amounts of fluid flowing in from each inlet. The interface is positioned between two adjacent streams — one with and the other without the molecule under investigation — over a single cell spread across the floor of the channel (Fig. 1b).

By streaming different solutions containing fluorescent tags over opposite poles of a live bovine capillary endothelial cell, we were able to maintain selective labelling of subpopulations of mitochondria in different regions of the cell (red and green in Fig. 1c, d) for over 2.5 hours, during which time the two subpopulations could be seen moving and intermixing inside the cell.

Figure 1e, f shows the disruption of actin filaments in selected regions of the cell after treatment with latrunculin A, a membrane-permeable molecule that binds to actin

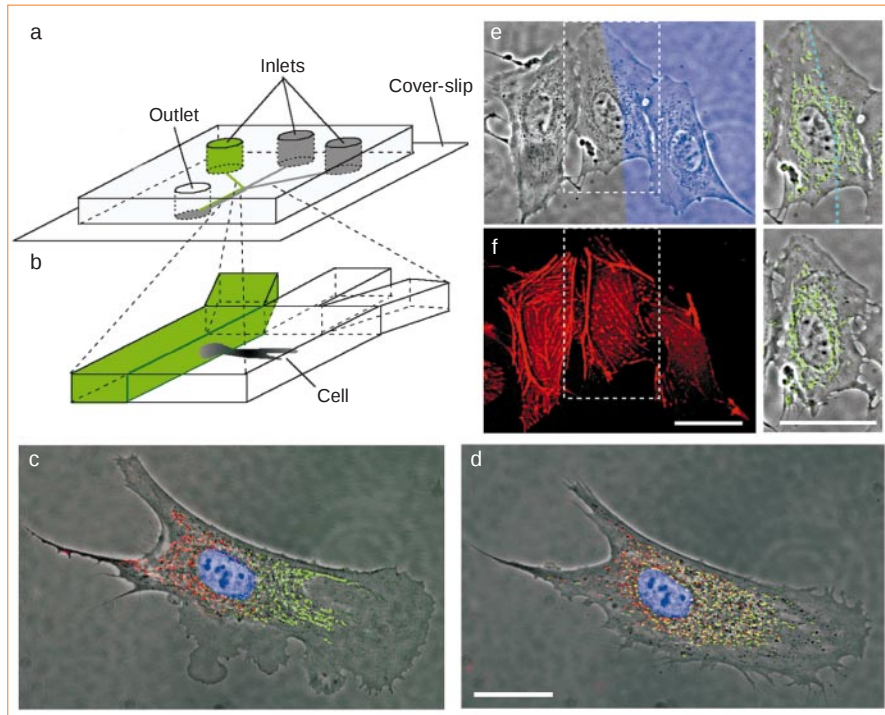


Figure 1 Differential manipulation of regions of a single bovine capillary endothelial cell using multiple laminar flows. **a**, **b**, Experimental set-up; **b** shows a close-up of the point at which the inlet channels combine into one main channel. **c**, Fluorescence images of a single cell after treatment of its right pole with Mitotracker Green FM and its left pole with Mitotracker Red CM-H₂XROS. The entire cell is treated with the DNA-binding dye Hoechst 33342. **d**, Image of the same cell as in **c** but taken 2.5 h later, showing intermixing of the red and green subpopulations of mitochondria. **e**, Treatment of a portion of a single cell with latrunculin A. The blue region (Dextran-70,000 Cascade Blue stain) reveals the flow of medium containing latrunculin A. Right, enlarged view of the middle cell and its mitochondria (green). **f**, Phalloidin-Alexa 594-labelled image of the same cell immediately after 10 min of treatment with latrunculin A. The cell in the middle, which was only partly in the stream containing latrunculin, shows disruption of actin microfilaments that is limited to this region (scale bars, 25 µm). Further details are available from the authors.

monomers and promotes the breaking up of polymeric actin filaments^{8,9}. The disruption evident on the right of the cell, which was induced by targeted localization of latrunculin A, causes the mitochondria and the nucleus to shift to the left, although the peripheral shape of the cell remains relatively unchanged (Fig. 1f). This displacement of mitochondria, which are mainly associated with microtubules¹⁰, by local disruption of actin microfilaments is consistent with the existence of mechanical connections between these two filament systems².

Our results indicate that partial treatment using laminar flows is an effective way to deliver small molecules to selected domains inside single mammalian cells. It is surprising that rapidly diffusing, membrane-permeable molecules can be directed so precisely, given the minute size of the cells themselves. Localization is achieved by a combination of rapid influx and efflux of molecules in distinct regions of the cell, created by using multiple laminar streams. Operation is straightforward and requires no special equipment apart from the moulds that form the capillary channels. Potential applications of our technique include the study of cell dynamics, chemotaxis, cell polarity, spatially regulated signalling, drug screening, and other

phenomena that involve intracellular compartments and subcellular heterogeneity.

Shuichi Takayama*†, Emanuele Ostuni*†, Philip LeDuc‡, Keiji Naruse†‡, Donald E. Ingber‡, George M. Whitesides*

*Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, USA

e-mail: gwhitesides@gmwhgroup.harvard.edu

‡Departments of Pathology and Surgery, Children's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA

†Present addresses: Department of Biomedical Engineering, University of Michigan, Ann Arbor, Michigan 48109-2125, USA (S.T.); Surface Logix Inc., Brighton, Massachusetts 02135, USA (E.O.); Department of Physiology, Nagoya University School of Medicine, Aichi 466-8550, Japan (K.N.)

- Huang, S. & Ingber, D. E. *Nature Cell Biol.* **1**, 131–138 (1999).
- Wang, N., Butler, J. P. & Ingber, D. E. *Science* **260**, 1124–1127 (1993).
- Duffy, D. C., McDonald, J. C., Schueller, O. J. A. & Whitesides, G. M. *Anal. Chem.* **70**, 4974–4984 (1998).
- Takayama, S. et al. *Proc. Natl Acad. Sci. USA* **96**, 5545–5548 (1999).
- Xia, Y. & Whitesides, G. M. *Angew. Chem. Int. Edn Engl.* **37**, 550–575 (1998).
- Wieg, B. H. & Yager, P. *Science* **283**, 346–347 (1999).
- Kenis, P. J. A., Ismagilov, R. F. & Whitesides, G. M. *Science* **285**, 83–85 (1999).
- Ayscough, K. *Methods Enzymol.* **298**, 18–25 (1998).
- Bradke, F. & Dotti, C. G. *Science* **283**, 1931–1934 (1999).
- Ball, E. H. & Singer, S. J. *Proc. Natl Acad. Sci. USA* **79**, 123–126 (1982).

Specific protection against breast cancers by cyclin D1 ablation

Qunyan Yu, Yan Geng & Piotr Sicinski

Department of Cancer Biology, Dana-Farber Cancer Institute, and Department of Pathology, Harvard Medical School, Boston, Massachusetts 02115, USA

Breast cancer is the most common malignancy among women. Most of these cancers overexpress cyclin D1, a component of the core cell-cycle machinery. We previously generated mice lacking cyclin D1 using gene targeting. Here we report that these cyclin D1-deficient mice are resistant to breast cancers induced by the *neu* and *ras* oncogenes. However, animals lacking cyclin D1 remain fully sensitive to other oncogenic pathways of the mammary epithelium, such as those driven by *c-myc* or *Wnt-1*. Our analyses revealed that, in mammary epithelial cells, the Neu–Ras pathway is connected to the cell-cycle machinery by cyclin D1, explaining the absolute dependency on cyclin D1 for malignant transformation in this tissue. Our results suggest that an anti-cyclin D1 therapy might be highly specific in treating human breast cancers with activated Neu–Ras pathways.

Cyclin D1 belongs to the family of three closely related D-type cyclins, termed cyclin D1, D2 and D3. These three proteins are expressed in an overlapping, redundant fashion in all proliferating cell types. D-cyclins collectively control cell-cycle progression by activating their cyclin-dependent kinase partners, CDK4 and CDK6, which leads to phosphorylation of the retinoblastoma protein, and in turn to the advance through the G1 phase of the cell cycle¹.

Several lines of evidence point to an important role for cyclin D1 in breast cancer formation. The *cyclin D1* gene is amplified in up to 20% of human breast cancers², while cyclin D1 protein is overexpressed in over 50% of human mammary carcinomas^{3–5}. The overexpression of cyclin D1 is seen in all histological types of human breast cancers³. It can be detected at the earliest stages of breast cancer progression, such as ductal carcinoma *in situ*, but not in premalignant lesions⁶. Once acquired, overexpression of cyclin D1 is maintained in all stages of the disease including the metastatic lesions^{3,7}. Importantly, overexpression of cyclin D1 seems to have a causative role in breast cancer formation, as transgenic mice engineered to overexpress cyclin D1 in mammary glands succumb to breast cancers⁸.

We and others previously generated cyclin D1-deficient mice^{9,10}. We found that these *cyclin D1*^{−/−} animals were viable and showed a narrow set of developmental abnormalities restricted to the retina and the nervous system^{9,10}. In adult mice, however, loss of cyclin D1 had virtually no impact on mouse physiology, except that the mammary glands of *cyclin D1*^{−/−} females failed to undergo full lobuloalveolar development during the late stage of pregnancy. This defect was restricted to pregnancy-associated proliferation, because *cyclin D1*^{−/−} mice developed normal mammary glands during sexual maturation^{9,10}.

The limited impact of cyclin D1 loss on mouse physiology—together with the well documented role of cyclin D1 overexpression in human breast cancers—suggested to us that the ablation of cyclin D1 might be highly selective in shutting off the proliferation of breast-tumour cells while sparing other tissues. As a first step towards a potential strategy for anti-cyclin D1 therapy in human breast cancers, we investigated whether the ablation of cyclin D1 protects *cyclin D1*^{−/−} mice against breast cancers.

To address this possibility, we crossed *cyclin D1*^{−/−} mice with four different strains of breast-cancer-prone mouse mammary tumour virus (MMTV)-oncogene transgenic mice, and we generated *cyclin D1*^{−/−}/MMTV-oncogene animals. For our studies we used strains overexpressing the oncogenes *v-Ha-ras* (ref. 11), *c-neu* (ref. 12), *c-myc* (ref. 13) and *Wnt-1* (ref. 14).

Analyses of mammary glands

We started our analyses by determining the expression pattern of D-cyclins in mammary glands of non-transgenic, virgin wild-type and *cyclin D1*^{−/−} females. Wild-type mammary glands expressed mostly cyclin D1, together with low levels of cyclin D2 and D3. Mammary glands of *cyclin D1*^{−/−} females lacked cyclin D1, but instead contained modestly elevated levels of cyclin D2 and slightly increased levels of cyclin D3 (Fig. 1c). We presume that these low levels of cyclins D2 and D3 allow normal mammary development in *cyclin D1*^{−/−} virgin mice.

We next compared the appearance of mammary glands of adult, virgin *cyclin D1*^{−/−}/MMTV-oncogene females with that of *cyclin D1*^{+/+}/MMTV-oncogene females. For each of four transgenic strains, we found that the appearance of *cyclin D1*^{−/−} mammary glands was identical to that of wild-type mice (Fig. 1a). This is consistent with our earlier observations that *cyclin D1*^{−/−} mice develop normal mammary glands in a virgin state⁹. For our tumour-susceptibility analyses, all females were kept as virgins throughout the entire observation period, except for the MMTV-*myc* mice (see Methods).

These control experiments provided us with an additional, unexpected observation. As reported previously¹⁴, mammary glands of MMTV-*Wnt-1* transgenic mice undergo precocious lobuloalveolar development in a virgin state. As a result, mammary glands of MMTV-*Wnt-1* virgin females (and males) resemble those of pregnant wild-type, non-transgenic females¹⁴. Strikingly, we observed the same phenotype in *cyclin D1*^{−/−}/MMTV-*Wnt-1* mice (Fig. 1a). This suggests that *Wnt-1*-dependent proliferative signals do not require cyclin D1. This is in contrast with recent reports that the *Wnt-1*– β -catenin signalling pathway critically impinges on cyclin D1 (refs 15, 16). It also reveals that *cyclin D1*^{−/−} mammary epithelium can undergo lobuloalveolar development under certain conditions.

Incidence of breast cancer

We observed MMTV-oncogene mice for breast cancer incidence. We found that the loss of cyclin D1 did not protect *cyclin D1*^{−/−} mice from breast cancers induced by the *myc* and *Wnt-1* oncogenes (Fig. 2 and Table 1). In marked contrast, *cyclin D1*^{−/−} mice were resistant to breast cancers induced by the *ras* and *neu* oncogenes. Thus, during the observation period, 19 of 21 *cyclin D1*^{+/+}/MMTV-*ras* mice died of breast cancers, developing a total of 39 tumours, whereas all 18 *cyclin D1*^{−/−}/MMTV-*ras* females remained free of tumours (Fig. 2 and Table 1). Likewise, 26 of 26 *cyclin D1*^{+/+}/MMTV-*neu* animals died of mammary carcinomas, developing a total of 79 tumours,

Table 1 Incidence of mammary carcinomas in transgenic mice

	<i>D1^{-/-}</i> /MMTV	<i>D1^{+/-}</i> /MMTV	<i>D1^{-/-}</i>	Wild type	<i>D2^{-/-}</i> /MMTV	<i>D3^{-/-}</i> /MMTV
MMTV- <i>v-Ha-ras</i>	0/18 (0)	19/21 (39)			6/11 (20)	8/13 (25)
MMTV- <i>c-neu</i>	0/42 (0)	26/26 (79)				
MMTV- <i>c-myc</i>	12/20 (18)	16/22 (21)				
MMTV- <i>Wnt-1</i>	15/16 (28)	20/20 (38)				
Non-transgenic			0/27 (0)	0/20 (0)		

The observation period was 70 weeks for MMTV-*c-neu*, MMTV-*c-myc* transgenic mice and for non-transgenic wild-type and *cyclin D1^{-/-}* controls, and 50 weeks for MMTV-*ras* and MMTV-*Wnt-1* animals. The ratios of females displaying mammary carcinomas to the total number of observed females are shown, along with total number of breast tumours observed among females of each genotype (in parentheses).

whereas all 42 *cyclin D1^{-/-}*/MMTV-*neu* mice remained healthy and free of tumours during the observation period (Fig. 2 and Table 1).

The possibility that these differences were caused by an inadequate expression of *ras* or *neu* transgenes in the mammary glands of *cyclin D1^{-/-}* animals was ruled out by the reverse transcription with polymerase chain reaction (PCR) analyses. These analyses revealed similar levels of the transgenes in the mammary glands of wild-type and *cyclin D1^{-/-}* females (Fig. 1b). We concluded that

cyclin D1 is critically required for Ras- and Neu-induced mammary tumorigenesis, and, consequently, the loss of *cyclin D1* renders *cyclin D1^{-/-}* mice resistant to breast cancers induced by these two oncogenes.

To determine whether this critical role is uniquely associated with *cyclin D1*, we crossed mice lacking either of the other two members of the D-cyclin family, namely *cyclin D2* (ref. 17) and *cyclin D3* (E. Sicinska and P.S., in preparation) with MMTV-*ras* mice. We found that *cyclin D2^{-/-}*/MMTV-*ras* and *cyclin D3^{-/-}*/MMTV-*ras* mice were susceptible to breast cancers (Table 1), pointing to a unique role for *cyclin D1* in Ras-induced breast tumorigenesis.

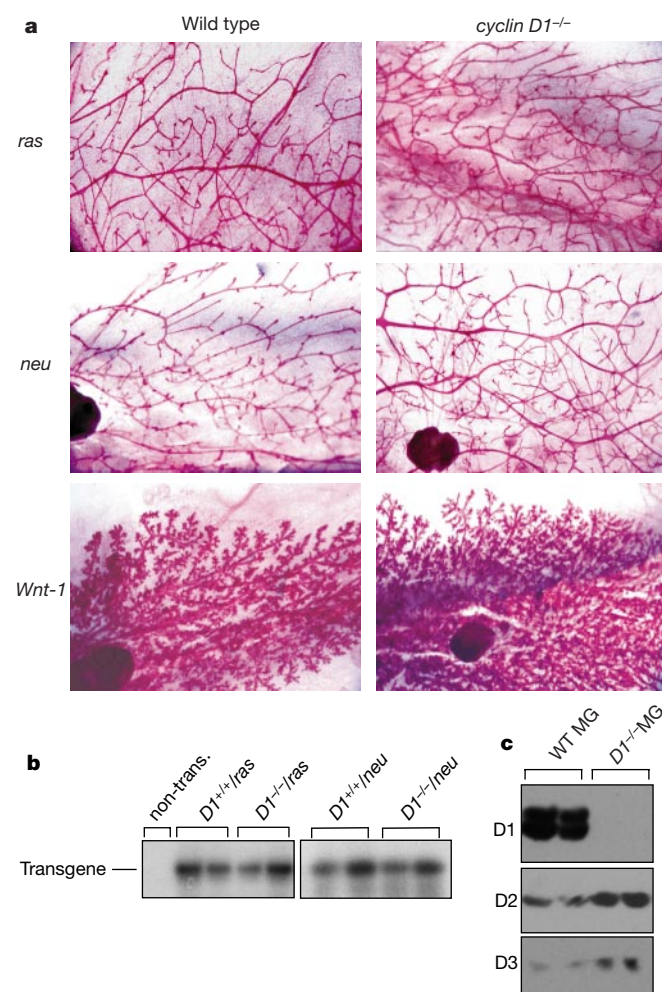


Figure 1 Analyses of mammary glands. **a**, Appearance of mammary glands in *cyclin D1* wild-type/MMTV-oncogene and *cyclin D1^{-/-}*/MMTV-oncogene virgin females. Mammary epithelial whole mounts were stained with carmine red. Magnification $\sim 3\times$.

b, Expression of the transgenes in wild-type (*cyclin D1^{+/+}*) and *cyclin D1^{-/-}* mammary glands (see Methods). non-trans., non-transgenic control; *D1^{+/+}/ras*, *cyclin D1^{+/+}*/MMTV-*v-Ha-ras*; *D1^{-/-}/ras*, *cyclin D1^{-/-}*/MMTV-*v-Ha-ras*; *D1^{+/+}/neu*, *cyclin D1^{+/+}*/MMTV-*c-neu*; *D1^{-/-}/neu*, *cyclin D1^{-/-}*/MMTV-*c-neu*. **c**, Expression of cyclins D1, D2 and D3 in mammary glands of virgin wild-type (WT MG) and *cyclin D1^{-/-}* female mice (*D1^{-/-}*-MG), as detected by western blotting.

Expression of D-cyclins in breast tumours

To understand the molecular basis of this strict requirement for *cyclin D1* in Ras- and Neu-induced, but not in Wnt-1- and Myc-driven, mammary tumorigenesis, we examined the expression pattern of the three D-type cyclins in breast tumours arising in *cyclin D1^{+/+}* transgenic females. These analyses revealed that breast tumours arising in MMTV-*ras* and MMTV-*neu* mice expressed virtually only *cyclin D1* (no *cyclin D2*, only very low levels of *cyclin D3*; Fig. 3a and c, first lane). In contrast, several tumours arising in MMTV-*Wnt-1* and MMTV-*myc* females expressed, in addition to *cyclin D1*, also high levels of *D2* (Fig. 3a). Importantly, we verified that all tumours included in analyses arose from luminal epithelial cells, as they expressed keratin 19 (data not shown). These findings indicate that, in mammary epithelial cells, *ras* and *neu* oncogenes communicate with the cell-cycle machinery through *cyclin D1*, whereas *Wnt-1* and *myc* can signal through other targets.

We extended these analyses by studying pure populations of mouse breast cancer cells grown *in vitro*. We compared the expression pattern of D-cyclins between a non-transformed mouse mammary epithelial cell line, HC11 (ref. 18), a breast-cancer cell line, SH1.1, derived from a tumour arising in a MMTV-*v-Ha-ras* female mouse, and a breast cancer cell line, 13Ma1a, derived from a tumour arising in a MMTV-*c-myc* female mouse (the two transgenic strains used as a source of tumour cells are the same as those used in our tumour-susceptibility analyses). As expected¹⁹, non-transformed mammary epithelial cells expressed *cyclin D1* as the major D-cyclin (Fig. 3b). *ras*-transformed breast cancer cells displayed grossly elevated levels of *cyclin D1*, but did not express *cyclin D2*. In contrast, *myc*-transformed breast cancer cells contained high levels of *cyclin D2* in addition to *cyclin D1* (Fig. 3b). Importantly, we verified that the Myc-expressing breast cancer cell line, 13Ma1a, arose from luminal epithelial cells, as it expressed keratin 19 (data not shown). The results of this experiment strongly support our interpretation that the *ras* oncogene (and the *neu* oncogene, which is known to operate upstream of *ras* (ref. 20) signals in mammary epithelial cells by inducing exclusively *cyclin D1*, whereas the *myc* oncogene can act through other targets in these cells.

Ras and Neu action in other cell types

We next asked whether the absolute requirement for *cyclin D1* in Ras- and Neu-driven tumorigenesis operated in all cell types of a body. Our analyses of MMTV-oncogene females indicated that this was not true. Thus, one of the four transgenic strains used for our analyses, that is MMTV-*ras* mice, succumb to salivary gland

tumours, albeit with a very low frequency, in addition to high-frequency mammary carcinomas¹¹. We observed one salivary carcinoma among 21 *cyclin D1*^{+/+}/MMTV-*ras* females (not shown), with 19 of the mice displaying breast tumours. We also observed three salivary adenocarcinomas among *cyclin D1*^{-/-}/MMTV-*ras* mice (not shown), while none of these mice developed breast tumours. Hence, these *cyclin D1*^{-/-}/MMTV-*ras* mice are susceptible to Ras-induced salivary tumours in the absence of cyclin D1.

To explore this idea further, we infected wild-type and *cyclin D1*^{-/-} 3T3 mouse embryo fibroblasts with retroviruses encoding activated *ras* or *neu* oncogenes, and scored the ability of these cells to grow in soft agar and to form tumours in nude mice. These analyses revealed that *cyclin D1*^{-/-} fibroblasts became fully transformed by *ras* and *neu* oncogenes, grew vigorously in soft agar (Fig. 4a) and formed tumours in nude mice (Fig. 4b). Hence, in contrast to mammary epithelial cells, in fibroblasts *ras* and *neu* oncogenes are able to drive malignant transformation in the absence of cyclin D1.

To understand this difference between mammary epithelial cells and fibroblasts, we examined the expression of D-cyclins in tumours formed by *ras*- and *neu*-transformed fibroblasts injected into nude mice, and compared it with the pattern observed in mammary epithelial tumours. As described above, mammary tumours induced by Ras or Neu expressed cyclin D1 as the major D-type cyclin (Fig. 3c, lane 1). In contrast, Ras- and Neu-driven fibroblast tumours expressed high levels of cyclin D2 and D3 in addition to cyclin D1 (Fig. 3c, lanes 2–6). We conclude that—unlike in mammary epithelial cells—*ras* and *neu* oncogenes in fibroblasts can communicate with the cell-cycle machinery also through cyclins D2 and D3. Consequently, *ras* and *neu* oncogenes continue to elicit malignant transformation even in the absence of cyclin D1. We presume that the same is true in other cell types where cyclin D1 was shown to be dispensable for transformation by *ras*. However, our analyses do not allow us to distinguish whether the induction itself, or simply the presence of particular combinations

of D-cyclins, is required for malignant transformation of different cell types by oncogenic Ras and Neu.

Discussion

The *ras* oncogene, acting through the mitogen-activated protein kinase pathway, is known to induce cyclin D1 by acting on cyclin D1 promoter^{21–24}. Similarly, the *neu* oncogene, which is known to operate upstream of *ras*, upregulates the promoter of the *cyclin D1* gene²⁵. Our analyses of *cyclin D1*^{-/-} mice revealed that the *ras* and *neu* oncogenes are absolutely dependent on cyclin D1 for malignant transformation of mammary glands. Moreover, we found that in mammary epithelial cells, activation of *ras* and *neu* oncogenes leads to induction of cyclin D1 messenger RNA, but not mRNA of cyclin D2 or cyclin D3. We conclude that in this particular cell type, *ras* and *neu* oncogenes are connected to the cell-cycle machinery exclusively through the promoter of the *cyclin D1* gene, explaining the absolute dependency of these two oncogenes on cyclin D1.

This conclusion is strengthened by our analyses of cyclin E→D1 ‘knock-in’ mice that we generated²⁶. In these knock-in animals, the coding sequences of the *cyclin D1* gene have been deleted and replaced by that of human *cyclin E*. In the tissues of cyclin E→D1 mice the expression of human cyclin E is driven by the cyclin D1 promoter, and it closely mimics the expression pattern of cyclin D1 in wild-type mice²⁶. We crossed cyclin E→D1 mice with an MMTV-*neu* strain and generated cyclin E→D1/MMTV-*neu* females. These females, lacking cyclin D1, displayed an incidence of breast cancer similar to that of *cyclin D1*^{+/+}/MMTV-*neu* animals (data not shown). Notably, we found that breast cancers arising in cyclin E→D1/MMTV-*neu* females expressed high levels of human cyclin E mRNA and protein (data not shown). These findings further confirm that the Neu–Ras pathway acts through regulatory elements located within the cyclin D1 promoter. They also reveal that cyclin E can replace cyclin D1 not only in mouse development—as was shown before²⁶—but also in driving proliferation of breast cancer cells.

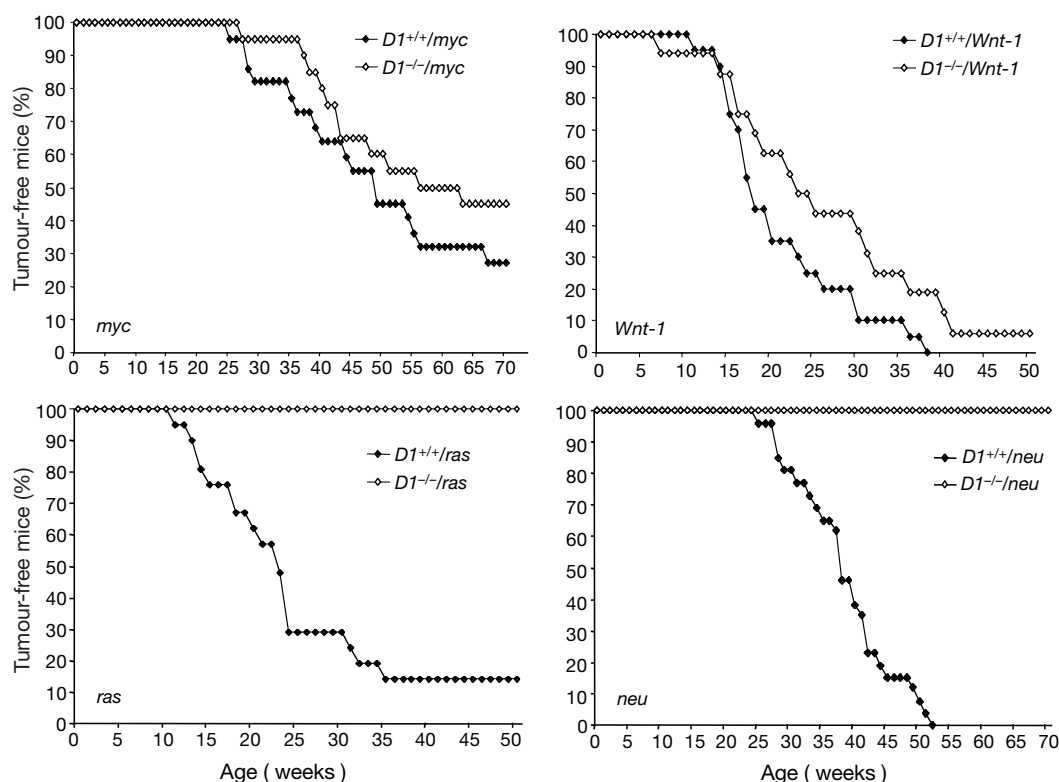


Figure 2 The occurrence of breast cancers in *cyclin D1*^{+/+} and *cyclin D1*^{-/-} transgenic mice. *myc*, MMTV-*c-myc* transgene; *Wnt-1*, MMTV-*Wnt-1* transgene, *ras*, MMTV-*v-Ha-ras* transgene, *neu*, MMTV-*c-neu* transgene.

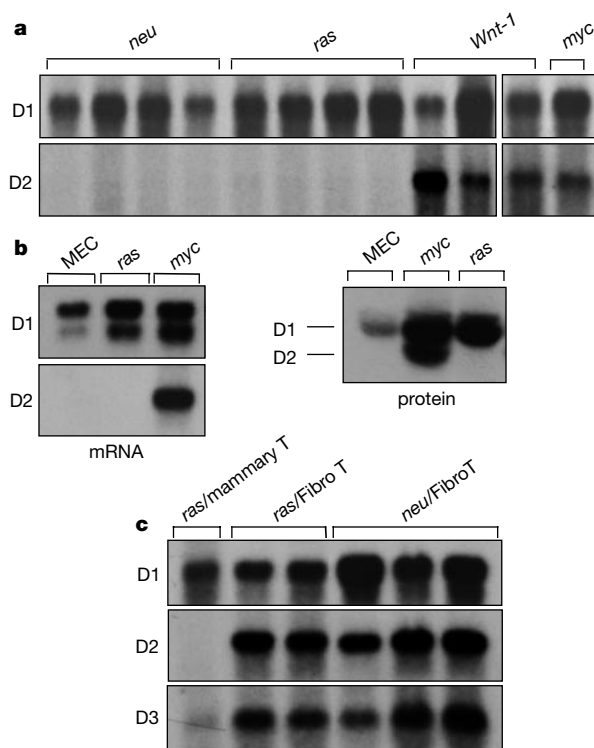


Figure 3 Levels of D-cyclins in tumours. **a**, Expression of cyclins D1 and D2 in tumours arising in MMTV-*c-neu*, MMTV-*v-Ha-ras*, MMTV-*Wnt-1* and MMTV-*c-myc* transgenic females, as detected by northern blotting. **b**, Expression of cyclins D1 and D2 in a non-transformed mouse mammary epithelial cell line, HC11 (MEC), in breast cancer cell line SH1.1 (derived from a MMTV-*v-Ha-ras* transgenic mouse) (*ras*), and in breast cancer cell line 13Ma1a (derived from a MMTV-*c-myc* transgenic mouse) (*myc*), as detected by northern blotting (left panel) and western blotting (right panel). **c**, Expression of D-cyclins in tumours formed by *ras*- (*ras*/Fibro T) and *neu*-transformed fibroblasts (*neu*/Fibro T) injected into nude mice, as detected by northern blotting. Result from a mammary tumour deriving from a MMTV-*v-Ha-ras* mouse (*ras*/Mammary T) is presented in lane 1 for comparison.

In contrast to Ras- and Neu-induced carcinogenesis, our analyses revealed that *myc* and *Wnt-1* oncogenes can communicate with the cell-cycle machinery in mammary epithelial cells through targets other than cyclin D1. Indeed, the *myc* oncogene may be connected to the cell-cycle machinery through targets acting downstream of D-cyclins^{27,28}. Consequently, Myc (as well as Wnt-1) continue to elicit malignant transformation even in the absence of cyclin D1.

In clear contrast to mammary epithelial cells, we found that in other cell types *ras* and *neu* oncogenes continued to elicit malignant transformation even in the absence of cyclin D1. Robles *et al.* showed that *cyclin D1*^{-/-} mice remain susceptible to Ras-driven skin papillomas, although the number of skin tumours per mouse was lower in *cyclin D1*^{-/-} mice than in *cyclin D1*^{+/+} animals²⁹. Our analyses suggest that, in cell types other than mammary epithelium, *ras* and *neu* oncogenes can also signal through cyclins D2 and D3. These results challenge the notion that the wiring of oncogenic pathways to the cell-cycle machinery operates in the same fashion in all cell types and indicate that this wiring is specific to cell type.

The unique requirement for cyclin D1 in Ras- and Neu-induced mammary tumorigenesis, demonstrated here, together with earlier observations that ablation of cyclin D1 had virtually no consequences on adult mice physiology^{9,10}, suggests that anti-cyclin D1 therapy might be highly selective in shutting off the growth of human breast cancers while sparing other tissues. About 50% of human breast cancers contain amplification and/or overexpression of the *c-neu* (*c-erbB-2*, *HER-2*) gene³⁰. Although mutations within

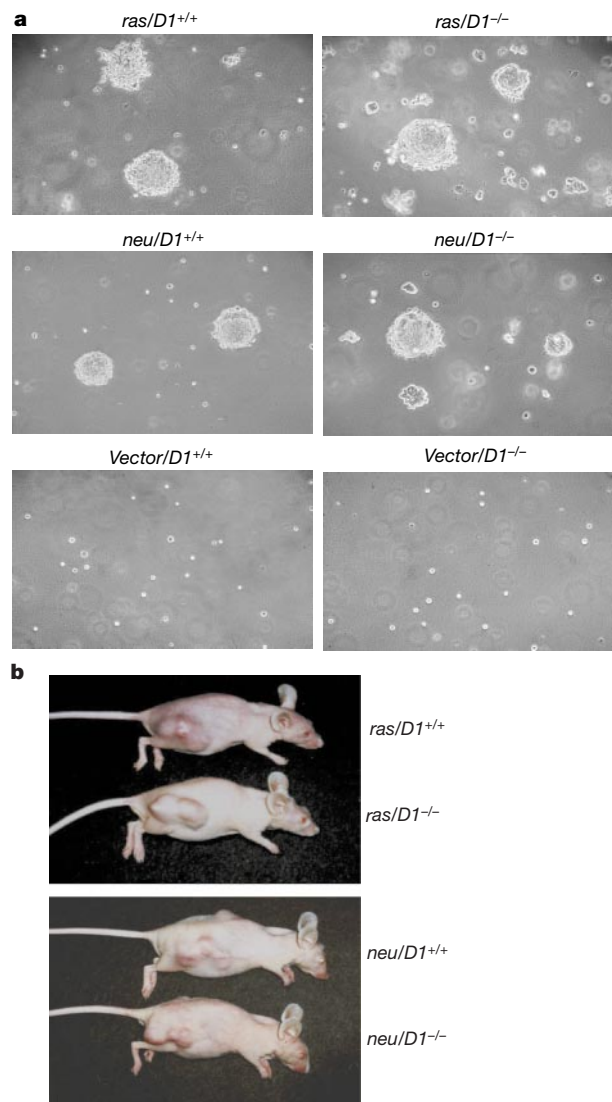


Figure 4 Malignant transformation of *cyclin D1*^{-/-} cells by Ras and Neu. **a**, Wild-type (*cyclin D1*^{+/+}) and *cyclin D1*^{-/-} 3T3 mouse embryo fibroblasts were infected with retroviruses encoding activated *ras* or *neu* oncogenes or with a control (empty) vector, and the ability of cells to form colonies in soft agar was scored. The photographs were taken after 13 days. Magnification ~15×. **b**, The cells from **a** were injected into nude mice, and the ability to form tumours was scored. Animals were photographed 3 weeks (*ras*) or 4 weeks (*neu*) after the injection. Cells infected with an empty vector were injected into the contralateral flanks and did not form tumours (not shown).

the *ras* gene in human breast tumours are quite rare, several oncogenic pathways—including those originating from c-Neu/c-ErbB-2 and other receptor tyrosine kinases—were shown to signal through the Ras protein²⁰. Our work suggests that human breast cancers with activated Neu–Ras pathways are prime candidates for anti-cyclin D1 therapy. □

Methods

Mice

MMTV-*v-Ha-ras* (ref. 11), MMTV-*c-neu* (strain TG.NK, ref. 12) and MMTV-*c-myc* (ref. 13) mice were purchased from the Charles River Laboratories; MMTV-*Wnt-1* mice¹⁴ were purchased from the Jackson Laboratory. Transgenic mice were crossed with *cyclin D1*^{+/+} animals. The resulting *cyclin D1*^{+/+}/MMTV-oncogene males were back crossed with *cyclin D1*^{+/+} females yielding all four experimental groups: *cyclin D1*^{+/+}/MMTV-oncogene, *cyclin D1*^{-/-}/MMTV-oncogene, non-transgenic *cyclin D1*^{+/+} and *cyclin D1*^{-/-} mice. Similar crosses were performed with *cyclin D2*^{+/+} and *cyclin D3*^{+/+} mice. Only females were used for subsequent analyses. All females except for MMTV-*c-myc* mice were kept as virgins

throughout the entire observation period. *cyclin D1*^{+/+}/MMTV-*c-myc* and *cyclin D1*^{-/-}/MMTV-*c-myc* females underwent two or three pregnancies. Pups were removed immediately after birth, to minimize the impact of breast-feeding (*cyclin D1*^{-/-} females do not nurse their pups^{8,10}). Mice were monitored by palpation twice-weekly for tumours.

Whole mounts of mammary glands

Inguinal mammary glands were removed and whole mounts were prepared as described²⁶.

Transgene levels

Total RNA was prepared from mammary glands²⁶ and reverse transcribed using ProSTAR Kit (Stratagene). The resulting complementary DNA was subjected to 18–30 rounds of PCR amplification using transgene-specific primers: 5'-GCAACAGTCTCAACATTTCAC-3' and 5'-TCGACGACAAAGCAACAG-3' in 1 × PCR buffer (Perkin Elmer) containing 3.0 mM MgCl₂. The amplification products (predicted size 365 base pairs) were resolved on agarose gels, transferred to MagnaGraph membranes (Osmonics) and probed with a radiolabelled, internal, transgene-specific oligonucleotide, 5'-GGAAAGTGAAGGA-TAAGTGA-3'. For a control we used RNA prepared from mammary glands of non-transgenic littermates, which showed no amplification signal. Each lane corresponded to pooled mammary glands collected from 4–6 animals of each genotype.

Northern blotting

Total RNA was isolated; 20 µg of RNA was resolved with 1% MOPS-formaldehyde gels, transferred to MagnaGraph membrane (Osmonics), and probed with radiolabelled riboprobes specific for mouse cyclin D1, D2 and D3, as described²⁶.

Western blotting

Protein lysates were prepared; 50–150 µg of proteins were separated with 10% SDS-PAGE and transferred to Immobilon-P membrane (Millipore). The immunoblots were probed with the M-20 antibody (Santa Cruz), which recognizes cyclin D1 and D2, and anti-cyclin D1 (H-295, Santa Cruz) and D3 (C-16, Santa Cruz). As secondary antibodies, peroxidase-conjugated IgG (Jackson ImmunoResearch) were used followed by enhanced chemiluminescence (ECL) detection (Amersham).

Mammary cell lines

SH1.1 and 13Ma1a mammary carcinoma cell lines provided by P. Leder were grown in DMEM plus 10% bovine serum. Mouse non-transformed mammary epithelial cell line HC11 (ref. 18) was cultured in RPMI 1640 plus 8% bovine serum, 10 µg ml⁻¹ epidermal growth factor and 5 µg ml⁻¹ insulin.

Cells and retroviral infections

Wild-type and *cyclin D1*^{-/-} mouse fibroblasts were prepared from day 13.5 embryos, as described²⁶, and the 3T3 cells were established.

pBabe-puro-Ras-V12 retrovirus, encoding activated Ras, was provided by R. A. Weinberg. pBabe-puro-NeuT retrovirus, encoding activated rat *neu* cDNA containing Val→Glu substitution at amino acid 664, was created by subcloning the *SalI*-*HindIII* 4.6 kb fragment of the pSVNeuT plasmid³¹ into pBabe-puro vector. Phoenix cells were transfected with pBabe-puro-Ras-V12 or pBabe-puro-NeuT or with an empty pBabe-puro vector with calcium phosphate precipitation. Two days after the transfection, Phoenix cell supernatants were used to infect *cyclin D1*^{+/+} or *cyclin D1*^{-/-} 3T3 cells. Cells were selected for six days in puromycin (4 µg ml⁻¹).

Soft agar and tumorigenicity assays

Soft agar assays were performed using standard procedures. Briefly, logarithmically growing cells (10⁴ and 10⁵) were plated as single-cell suspensions in 0.4% agarose in DMEM supplemented with 10% fetal bovine serum. Cultures were analysed and scored after 5–28 days. Each cell line was analysed in quadruplicate. For tumorigenicity assays, 1 × 10⁶ and 5 × 10⁶ cells, resuspended in 200 µl PBS, were injected subcutaneously into flanks of female 6–8-week-old CD-1 nude mice (Charles River Laboratories). Mice were monitored for tumour formation and were killed 21–28 days after the injection. Eight mice were used to evaluate the tumorigenicity of a given cell line.

Received 23 March; accepted 9 May 2001.

- Sherr, C. J. & Roberts, J. M. CDK inhibitors: positive and negative regulators of G1-phase progression. *Genes Dev.* **13**, 1501–1512 (1999).
- Dickson, C. *et al.* Amplification of chromosome band 11q13 and a role for cyclin D1 in human breast cancer. *Cancer Lett.* **90**, 43–50 (1995).

- Bartkova, J. *et al.* Cyclin D1 protein expression and function in human breast cancer. *Int. J. Cancer* **57**, 353–361 (1994).
- Gillett, C. *et al.* Amplification and overexpression of cyclin D1 in breast cancer detected by immunohistochemical staining. *Cancer Res.* **54**, 1812–1817 (1994).
- McIntosh, G. G. *et al.* Determination of the prognostic value of cyclin D1 overexpression in breast cancer. *Oncogene* **11**, 885–891 (1995).
- Weinstat-Saslow, D. *et al.* Overexpression of cyclin D mRNA distinguishes invasive and *in situ* breast carcinomas from non-malignant lesions. *Nature Med.* **1**, 1257–1260 (1995).
- Gillett, C. *et al.* Cyclin D1 and prognosis in human breast cancer. *Int. J. Cancer* **69**, 92–99 (1996).
- Wang, T. C. *et al.* Mammary hyperplasia and carcinoma in MMTV-cyclin D1 transgenic mice. *Nature* **369**, 669–671 (1994).
- Sicinski, P. *et al.* Cyclin D1 provides a link between development and oncogenesis in the retina and breast. *Cell* **82**, 621–630 (1995).
- Fantl, V., Stamp, G., Andrews, A., Rosewell, I. & Dickson, C. Mice lacking cyclin D1 are small and show defects in eye and mammary gland development. *Genes Dev.* **9**, 2364–2372 (1995).
- Sinn, E. *et al.* Coexpression of MMTV/*v-Ha-ras* and MMTV/*c-myc* genes in transgenic mice: synergistic action of oncogenes in vivo. *Cell* **49**, 465–475 (1987).
- Muller, W. J., Sinn, E., Pattengale, P. K., Wallace, R. & Leder, P. Single-step induction of mammary adenocarcinoma in transgenic mice bearing the activated *c-neu* oncogene. *Cell* **54**, 105–115 (1988).
- Stewart, T. A., Pattengale, P. K. & Leder, P. Spontaneous mammary adenocarcinomas in transgenic mice that carry and express MTV/*myc* fusion genes. *Cell* **38**, 627–637 (1984).
- Tsukamoto, A. S., Grosschedl, R., Guzman, R. C., Parslow, T. & Varmus, H. E. Expression of the *int-1* gene in transgenic mice is associated with mammary gland hyperplasia and adenocarcinomas in male and female mice. *Cell* **55**, 619–625 (1988).
- Rimerman, R. A., Gellert-Randleman, A. & Diehl, J. A. Wnt1 and MEK1 cooperate to promote cyclin D1 accumulation and cellular transformation. *J. Biol. Chem.* **275**, 14736–14742 (2000).
- Tetsu, O. & McCormick, F. Beta-catenin regulates expression of cyclin D1 in colon carcinoma cells. *Nature* **398**, 422–426 (1999).
- Sicinski, P. *et al.* Cyclin D2 is an FSH-responsive gene involved in gonadal cell proliferation and oncogenesis. *Nature* **384**, 470–474 (1996).
- Ball, R. K., Friis, R. R., Schoenenberger, C. A., Doppler, W. & Groner, B. Prolactin regulation of beta-casein gene expression and of a cytosolic 120-kd protein in a cloned mouse mammary epithelial cell line. *EMBO J.* **7**, 2089–2095 (1988).
- Lukas, J. *et al.* Cyclin D2 is a moderately oscillating nucleoprotein required for G1 phase progression in specific cell types. *Oncogene* **10**, 2125–2134 (1995).
- Janes, P. W., Daly, R. J., de Fazio, A. & Sutherland, R. L. Activation of the Ras signaling pathway in human breast cancer cells overexpressing *erbB-2*. *Oncogene* **9**, 3601–3608 (1994).
- Filmus, J. *et al.* Induction of Cyclin D1 overexpression by activated *ras*. *Oncogene* **9**, 3627–3633 (1994).
- Albanese, C. *et al.* Transforming p21ras mutants and *c-Ets-2* activate the cyclin D1 promoter through distinguishable regions. *J. Biol. Chem.* **270**, 23589–23597 (1995).
- Liu, J. J. *et al.* Ras transformation results in an elevated level of cyclin D1 and acceleration of G1 progression in NIH 3T3 cells. *Mol. Cell. Biol.* **15**, 3654–3663 (1995).
- Lavoie, J. N., L'Allemand, G., Brunet, A., Mueller, R. & Pouyssegur, J. Cyclin D1 expression is regulated positively by the p42/p44MAPK and negatively by the p38/HOGMAPK pathway. *J. Biol. Chem.* **271**, 20608–20616 (1996).
- Lee, R. J. *et al.* Cyclin D1 is required for transformation by activated Neu and is induced through an E2F-dependent signaling pathway. *Mol. Cell. Biol.* **20**, 672–683 (2000).
- Geng, Y. *et al.* Rescue of cyclin D1 deficiency by knockin cyclin E. *Cell* **97**, 767–777 (1999).
- Alevizopoulos, K., Vlach, J., Hennecke, S. & Amati, B. Cyclin E and *c-Myc* promote cell proliferation in the presence of p16INK4a and of hypophosphorylated retinoblastoma family proteins. *EMBO J.* **16**, 5322–5333 (1997).
- Santoni-Rugiu, E., Falck, J., Mailand, N., Bartek, J. & Lukas, J. Involvement of Myc activity in a G1/S-promoting mechanism parallel to the pRb/E2F pathway. *Mol. Cell. Biol.* **20**, 3497–3509 (2000).
- Robles, A. I. *et al.* Reduced skin tumor development in cyclin D1-deficient mice highlights the oncogenic *ras* pathway *in vivo*. *Genes Dev.* **12**, 2469–2474 (1998).
- Bieche, I. & Lidereau, R. Genetic alterations in breast cancer. *Genes Chromosom. Cancer* **14**, 227–251 (1995).
- Bargmann, C. I., Hung, M. C. & Weinberg, R. A. Multiple independent activations of the *neu* oncogene by a point mutation altering the transmembrane domain of p185. *Cell* **45**, 649–657 (1986).

Acknowledgements

We thank R. Bronson, M. Ciemerych, L. Le Cam, B. Elenbaas, A. Leder, S. Jean, M. Lee, M. Mann, T. Mäkelä, B. Rollins and R. A. Weinberg for help; P. Leder for SH1.1 and 13Ma1a cell lines; J. Drenzo and M. Brown for the HC11 cell line; and R. A. Weinberg for pBabe-puro-Ras-V12 and pSVNeuT plasmids. This work was supported by a grant from NIH and by the DOD Breast Cancer Idea Award to P.S. Y.G. is partly supported by the NCI SPOR in Breast Cancer at the DFCI/Harvard Medical School. P.S. is a Barr New Investigator.

Correspondence and requests for materials should be addressed to P.S. (e-mail: peter_sicinski@dfci.harvard.edu).

Gravitational microlensing by low-mass objects in the globular cluster M22

Kailash C. Sahu, Stefano Casertano, Mario Livio,
Ronald L. Gilliland, Nino Panagia, Michael D. Albrow & Mike Potter

Space Telescope Science Institute, 3700 San Martin Drive, Baltimore,
Maryland 21218, USA

Gravitational microlensing offers a means of determining directly the masses of objects ranging from planets to stars, provided that the distances and motions of the lenses and sources can be determined^{1,2}. A globular cluster observed against the dense stellar field of the Galactic bulge presents ideal conditions for such observations because the probability of lensing is high³ and the distances and kinematics of the lenses and sources are well constrained. The abundance of low-mass objects in a globular cluster is of particular interest, because it may be representative of the very early stages of star formation in the Universe, and therefore indicative of the amount of dark baryonic matter in such clusters. Here we report a microlensing event associated with the globular cluster M22. We determine the mass of the lens to be $0.13^{+0.03}_{-0.02}$ solar masses. We have also detected six events that are unresolved in time. If these are also microlensing events, they imply that a non-negligible fraction of the cluster mass resides in the form of free-floating planetary-mass objects.

The abundance of low-mass compact objects, at or below the hydrogen-burning mass limit of $0.08M_{\odot}$ (where $M_{\odot}$ is the solar mass), remains an open question. This abundance has implications for such diverse areas of astrophysics as the formation of stars and planets, the stellar initial mass function, the stellar mass–luminosity relation, and the properties of baryonic dark matter.

Globular clusters are of special interest in this context. Deep observations with the Hubble Space Telescope (HST) indicate that the stellar luminosity function in globular clusters drops rapidly at I-band magnitudes fainter than $M_I \approx 9$ (refs. 4–7). However, even HST observations cannot yet reach down to brown-dwarf (mass $\leq 0.08M_{\odot}$) luminosities in globular clusters, and the interpretation of the observed luminosity function in terms of a mass function, particularly near the hydrogen-burning limit, is complicated by the poorly known and metallicity-dependent mass–luminosity relation⁸. Another question is whether globular clusters contain planetary-mass objects. A recent attempt to detect planetary companions in the globular cluster 47 Tucanae by transits resulted in a non-detection⁹, implying that less than 0.1% of the stars in the cluster have Jupiter-size planets at orbital radii ≤ 0.05 astronomical units (AU); in contrast, $\sim 1\%$ of the stars in the solar neighbourhood have comparable planets¹⁰. If the scarcity of planetary companions in a globular cluster is due to environmental effects—such as an early disruption of planetary orbits caused by the high density of stars—planets could be expected to become detached from their parent stars, and hence not be detectable by a search for transits.

Detection and mass determination through microlensing do not suffer from the limitations affecting the above techniques, and offer two unique advantages. First, because microlensing is purely a gravitational effect, it is independent of the mass–luminosity–metallicity relation, cooling curve, or atmospheric properties. Second, as the light from the microlensing star need not be detected directly, microlensing is sensitive down to planetary masses. Gravitational lensing currently offers the only direct method for the determination of the low end of the mass function in globular clusters. Other searches for microlensing events have been per-

formed to find compact objects in the Galactic disk and halo, by looking towards the Galactic bulge and the Magellanic Clouds^{11–14}. Such programmes require the monitoring of a very large number of stars over areas of several square degrees, and the determination of the lens mass is uncertain because of the unknown distance and velocity of the lensing objects.

In contrast, our observations are directed towards the globular cluster M22, located about one-third of the way between the Sun and the Galactic bulge (distance from Earth, 2.6 kpc; galactic latitude and longitude, -7.6 and 9.9 degrees, respectively). The microlensing optical depth, which is the probability that a single background source is microlensed at any given time, is $\sim 10^{-5}/\phi$ at an angular distance ϕ from the cluster centre, where ϕ is expressed in arcmin (ref. 3). Thus, in the central region of the cluster, the microlensing optical depth is at least an order of magnitude higher than for the typical lines of sight explored in traditional microlensing survey programmes. The high spatial resolution of the HST and the richness of the background stellar field permit the monitoring of a large number of stars in the central region of M22, so that the detection of microlensing events is feasible with a moderate expenditure of HST time. The location of both potential lenses and sources and their relative proper motions are well constrained, removing the ambiguities usually present in the microlensing events detected towards the bulge and the Magellanic Clouds^{11–12,15–17}. The duration of a microlensing event (in days) can be used to calculate the mass of the lensing star, $t_{\text{ML}} \approx 49 (M_{\text{lens}}/1 M_{\odot})^{1/2}$, assuming a proper motion for M22 of 10.9 mas yr^{-1} (ref. 18), and distances of 2.6 kpc to M22¹⁹ and 8.2 kpc for a typical bulge star. Any time-resolved microlensing event thus provides a robust measurement of the mass of a lens in M22.

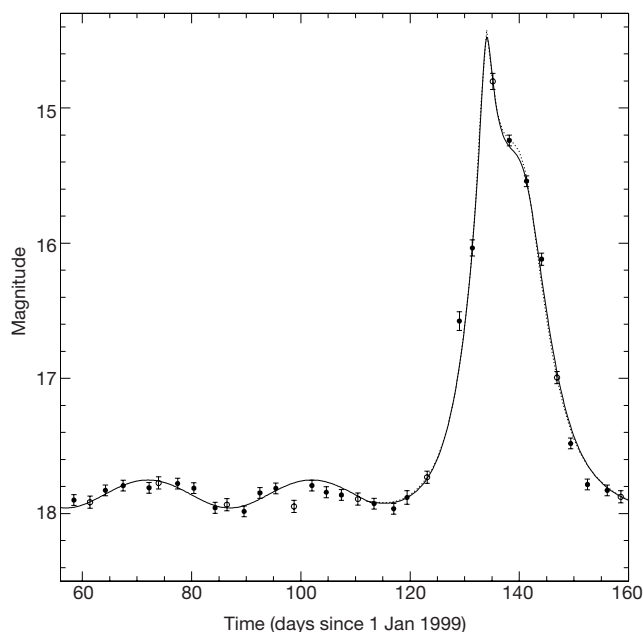


Figure 1 The observed light curve of the time-resolved microlensing event in the I and V bands, and the two best-fitting calculated microlensing light curves. The V magnitudes (open circles) are offset by -1.01 mag to show the amplifications on the same scale as the I magnitudes (filled circles). The light curve shows an additional periodic modulation with an amplitude of 0.19 mag and a periodicity of ~ 30 days which extends beyond the microlensing event. We attribute this periodic modulation to the source being a binary. The solid line corresponds to a binary in which the secondary component has no contribution to the total light, and the dotted line corresponds to a binary in which the secondary component contributes 10% to the total light. The microlensing timescale is ~ 17.6 days in either case, which corresponds to a lens mass of $0.13^{+0.03}_{-0.02} M_{\odot}$.

Our observational programme was performed from 22 February to 15 June 1999, using the WFPC2 camera on board HST. At each of 43 epochs, WFPC2 observations were taken of three fields in the central region of M22, oriented so as to maximize the observational efficiency. The images were taken with a typical separation of about 3 days. One additional image was taken a year later, on 18 February 2000, to check for possible variability of the sources between the different epochs. We monitored a total of $\sim 83,000$ bulge stars. Most of the observations were taken in the I (F814W) filter, with every fourth observation in the wide-V (F606W) filter to test for colour variations of the sources. We used the profile-fitting routines ('allstar') of the crowded field photometry package 'daophot'²⁰ to obtain photometric time series for all source stars. We then identified any stars that exhibited significantly larger-than-expected light variations within the time series, and inspected their light curves in greater detail.

Using this procedure, we detected one well-resolved event, the light curve of which is shown in Fig. 1. The event is achromatic—the I-based measurements (dots) and the V-band measurements (circles) lie on the same curve—and the light curve is very similar to that of an ideal microlensing event. The source star is clearly redder than the cluster main sequence at the same magnitude, and its properties are consistent with those of an F subgiant in the bulge.

A more detailed analysis of the light curve for this source shows a 30-day periodic modulation with an amplitude of 0.19 mag, extending beyond the microlensing event. We attribute this periodic modulation to the source being a binary. We then adopt a realistic model that includes source rotation, binary parameters, and light blending in a self-consistent way^{21,22}. The exact value of the binary's orbital separation has no significant consequence for our discussion. The two best-fitting solutions—both of which fit the observed data very well, and have a reduced χ^2 of 1.1—are shown in Fig. 1. The solid line corresponds to a binary in which the secondary

component has no contribution to the total light, and the dotted line corresponds to a binary in which the secondary component contributes 10% to the total light. Both solutions have a minimum impact parameter of 0.05 and a timescale of $t_E = 17.6$ days. We note that only about 15% of the bulge sources are brighter than 18th magnitude in I, and the fraction is smaller if we consider only the sources that show variability. The probability of one such star being microlensed is consistent with the estimated optical depth. Even so, the event seems slightly unusual, although such an *a posteriori* probability estimate is not statistically meaningful.

Because the probability of microlensing by a globular cluster star is much larger than that by a bulge star, the lens is assumed to be within the cluster and therefore its distance and kinematics are assumed known. The position and kinematics of the source are less well constrained, and cause some uncertainty in the derived lens mass. Based on the density distribution of the bulge stars along the line of sight and the luminosity of the source, the estimated distance to the source is 8.2 ± 1.0 kpc. The distance uncertainty of ± 1 kpc corresponds to a mass uncertainty of only $\sim 5\%$. The kinematics are less constrained because of the high velocity dispersion of the bulge objects²³. Taking this dispersion into account, the proper motion of M22 can be expressed as $10.9_{-0.9}^{+1.4}$ mas yr⁻¹, and the resultant mass of the lens is $0.13_{-0.02}^{+0.03} M_\odot$. Both uncertainties, and especially that associated with the kinematics, could be reduced by follow-up observations after a few years, when HST observations may enable a direct measurement of the proper motion of the source with respect to the globular cluster. The mass of the lens is consistent with the peak of the mass distribution of stars as estimated from the stellar luminosity function of other globular clusters^{9,10}.

Could the observed event be interpreted as some other type of stellar luminosity variation? The shape and achromaticity of the light curve argue in favour of microlensing, but we have nevertheless explored other phenomena that could explain a brightening of about 3 mag over a timescale of ~ 10 days. The most likely candidates are cataclysmic variables, such as classical novae and dwarf novae. Classical novae are excluded because of the relatively low amplitude of luminosity variation²⁴ and the lack of colour variation²⁵. Cataclysmic variables in general, with the exception of a few (such as BV Cen), are intrinsically blue, and none are known to have an orbital period longer than 6 days (ref. 26). Furthermore, in a diagram of timescale versus maximum amplitude, dwarf novae and microlensing events occupy distinct regions, and this event clearly falls in the region occupied by the microlensing events. In addition, several known variables of comparable brightness have been catalogued in this region of the sky, but none corresponds to this particular source²⁸. Although we cannot entirely rule out an unusual stellar variation as the cause of the event, this is very improbable, and the event is most naturally explained as being due to microlensing. Further photometric monitoring and spectroscopic and proper-motion observations of the source may help in confirming the microlensing nature of this event.

Although a single mass measurement does not by itself yield significant constraints on the mass function of M22, we have now shown that such observations are possible. It is very likely that, with a sufficient number of observations of M22 and other suitable

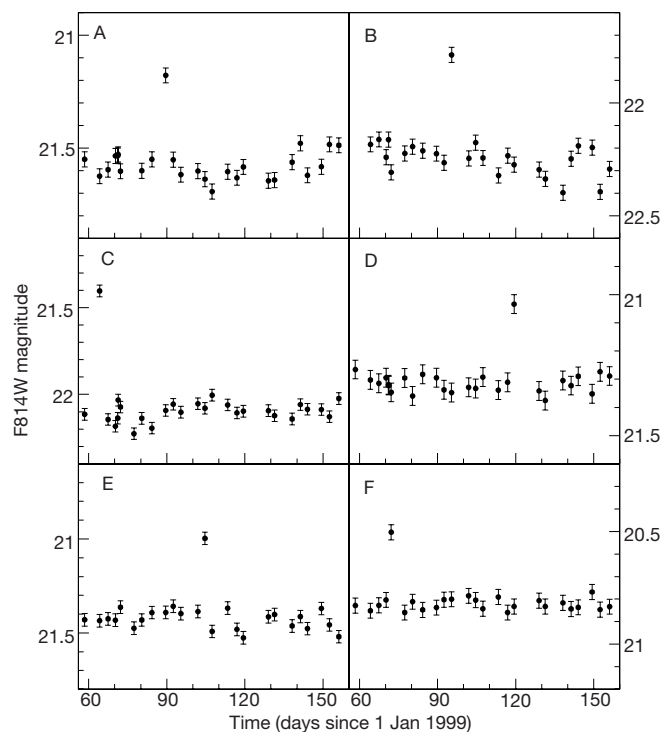


Figure 2 The unresolved microlensing candidate events. The photometric errors shown here, which are different for different events and which also depend on whether the source is amplified, were estimated by comparing the time-series magnitudes of stars of similar brightness in similar crowding conditions. See Table 1 for the sources of events A–F.

Table 1 Positions and magnitudes of the microlensed sources

Source*	RA (2000)	Dec. (2000)	m_{F814W}	m_{F606W}
Fig. 1	18:36:24.66	–23:54:35.5	17.85	18.86
A	18:36:28.63	–23:53:55.6	21.64	23.32
B	18:36:26.72	–23:54:18.8	22.25	23.81
C	18:36:31.24	–23:53:12.5	22.12	23.51
D	18:36:28.14	–23:52:57.5	21.33	22.70
E	18:36:24.09	–23:55:51.9	21.49	23.03
F	18:36:26.07	–23:53:36.6	20.81	22.14

* 'Fig. 1' refers to the microlensed source for the event shown in Fig. 1. A–F refer to the sources for the events shown in Fig. 2.

clusters, the shape of the low end of the mass distribution in globular clusters will become amenable to direct measurement.

We have also detected six events that could indicate a hitherto unknown population of planetary-mass objects in M22. Our time series shows six brightening events that are completely unresolved in time (Fig. 2). In each event, the star is brighter than its average by ~ 0.3 – 0.8 mag at one epoch, and returns to normal brightness at the next epoch. The observations of each field were taken in pairs, separated by about 6 minutes. In each case, the brightening is observed in both images, as required by the microlensing interpretation—microlensing of a bulge star cannot be shorter than about 2 hours because of the finite size of the source. We have carefully inspected the point-spread functions in each image, and have ruled out artefacts and imperfections in the data reduction process—such as cosmic-ray hits or bad pixels—as the cause of the brightening. The positions and the baseline magnitudes of all the microlensed sources are given in Table 1.

The interpretation of these events as microlensing is necessarily tentative, because we do not have the usual confirmation tools of the shape and achromaticity of the light curve. On the other hand, we know of no stellar variability that could explain our observations. We can rule out two possibilities: flare stars and cataclysmic variables. Flare stars are generally M dwarfs²⁹, which—depending on the distance—would be either too faint or too red for the observed magnitudes and colours of these stars. Cataclysmic variables are generally blue, whereas all our candidates are red. On the other hand, the luminosity and amplitude distributions of the candidates are precisely those expected from microlensing of bulge stars. Their number distribution follows the distribution of the bulge stars, suggesting that they are bulge sources, and their amplitude distribution is consistent with their being due to microlensing. The contribution of these six unresolved events to the total microlensing optical depth is small, and their combined microlensing optical depth is consistent with the expected optical depth.

If these events indeed represent gravitational microlensing, the statistical upper limit to their duration (at a 95% confidence level) is 0.8 days, corresponding to a mass of only ~ 0.25 times that of Jupiter. These objects must be either free-floating, or at least several astronomical units from any stellar-mass object, otherwise the effect of the stellar object on the light curve would be easily detectable. The total contribution of these free-floating planets (if real) to the cluster mass is uncertain, but taking into account the area and the sensitivity of our monitoring programme, we estimate that this is of the order of 10%. Future multi-band observations at high time resolution could unambiguously establish the nature of such events, and would determine if they are indeed caused by detached planetary-mass objects. □

Received 28 March; accepted 2 May 2001.

- Paczynski, B. Gravitational microlensing in the Local Group. *Annu. Rev. Astron. Astrophys.* **34**, 419–459 (1996).
- Schneider, P., Ehlers, J. & Falco, E. E. *Gravitational Lenses* (Springer, Berlin, 1993).
- Paczynski, B. Gravitational Microlensing by the Globular Cluster Stars. *Acta Astronomica* **44**, 235–239 (1994).
- Paresce, F., De Marchi, G. & Romaniello, M. Very low mass stars and white dwarfs in NGC 6397. *Astrophys. J.* **440**, 216–226 (1995).
- Elson, R. A. W., Gilmore, G. F., Santiago, B. X. & Casertano, S. HST observations of the stellar population of the globular cluster Ω Cen. *Astron. J.* **110**, 682–692 (1995).
- King, I., Anderson, J., Cool, A. M. & Piotto, G. The luminosity function of the globular cluster NGC 6397 near the limit of hydrogen burning. *Astrophys. J.* **492**, L37–L40 (1998).
- De Marchi, G. & Paresce, F. Low mass stars in globular clusters II. The mass function of M15. *Astron. Astrophys.* **304**, 202–210 (1995).
- Burrows, A., Hubbard, W. B., Saumon, D. & Lunine, J. I. An expanded set of brown dwarf and very low mass star models. *Astrophys. J.* **406**, 158–171 (1993).
- Gilliland, R. L. et al. A lack of planets in 47 Tucanae from a Hubble Space Telescope search. *Astrophys. J.* **545**, L47–L51 (2000).
- Marcy, G. W., Butler, R. P. Planets orbiting other suns. *Publ. Astron. Soc. Pacif.* **112**, 137–140 (2000).
- Alcock, C. et al. The MACHO project: microlensing results from 5.7 years of Large Magellanic Cloud observations. *Astrophys. J.* **542**, 281–307 (2000).
- Afonso, C. et al. Microlensing towards the Small Magellanic Cloud EROS 2 two-year analysis. *Astron. Astrophys.* **344**, L63–L66 (1999).

- Udalski, A. et al. The optical gravitational lensing experiment. Catalog of microlensing events in the galactic bulge. *Acta Astron.* **50**, 1–65 (2000).
- Yanagisawa, T. et al. Wide-field camera for gravitational microlensing survey: MOA-cam2. *Exp. Astron.* **10**, 519–535 (2000).
- Sahu, K. C. Stars within the Large Magellanic Cloud as potential lenses for observed microlensing events. *Nature* **370**, 275–276 (1994).
- Sahu, K. C. Microlensing events of the LMC are better explained by stars within the LMC than by MACHOs. *Publ. Astron. Soc. Pacif.* **106**, 942–948 (1994).
- Sahu, K. C. & Sahu, M. S. Spectroscopy of MACHO 97-SMC-1: Self-lensing within the Small Magellanic Cloud. *Astrophys. J.* **508**, L147–L150 (1998).
- Cudworth, K. M. & Hansen, R. B. Space velocities of 14 globular clusters. *Astron. J.* **105**, 168–172 (1983).
- Peterson, R. C. & Cudworth, K. M. Proper motions and radial velocities in the globular cluster M22 and the cluster distance. *Astrophys. J.* **420**, 612–631 (1994).
- Stetson, P. B. In *Stellar Photometry—Current Techniques and Future Developments* (eds Butler, C. J. & Elliot, I.) 291–303 (IAU Colloq. 136, Cambridge Univ. Press, Cambridge, 1993).
- Dominik, M. Galactic microlensing with rotating binaries. *Astron. Astrophys.* **329**, 361–374 (1998).
- Paczynski, B. P. Binary source parallactic effect in gravitational micro-lensing. Preprint astro-ph/9711007 at (<http://xxx.lanl.gov>) (1997).
- Freeman, K. C., de Vaucouleurs, G., Wainscoat, R. J. & de Vaucouleurs, A. A spectroscopic determination of the velocity dispersion in the Galactic bulge. *Astrophys. J.* **325**, 563–565 (1988).
- Bode, M. F. & Evans, A. E. *Classical Novae 5* (Wiley & Sons, Chichester, 1989).
- Warner, B. *Cataclysmic variable Stars 148* (Cambridge Univ. Press, Cambridge, 1995).
- Downes, R. et al. A catalog and atlas of cataclysmic variables—the living edition. *Publ. Astron. Soc. Pacif.* (in the press).
- Della Valle, M. & Livio, M. Are microlensing events contaminated by dwarf nova eruptions? *Astrophys. J.* **457**, L77–L79 (1996).
- Samus, N., Kravtsov, V., Pavlov, M., Alcaïno, G. & Liller, W. A new photographic colour-magnitude study of M 22 = NGC 6656. *Astron. Astrophys. Suppl. Ser.* **109**, 487–499 (1995).
- Petersen, B. R. A review of stellar flares and their characteristics. *Sol. Phys.* **121**, 299–312 (1989).

Acknowledgements

N.P. is on assignment from the Space Science Department of ESA. We thank H. Duerbeck, J. Valenti, H. Bond, R. Downes and J. Pringle for discussions.

Correspondence and requests for materials should be addressed to K.C.S. (e-mail: ksahu@stsci.edu).

Sub-poissonian loading of single atoms in a microscopic dipole trap

Nicolas Schlosser, Georges Reymond, Igor Protchenko & Philippe Grangier

Laboratoire Charles Fabry de l'Institut d'Optique, UMR 8501 du CNRS, BP 147, F91403 Orsay Cedex, France

The ability to manipulate individual atoms, ions or photons allows controlled engineering of the quantum state of small sets of trapped particles; this is necessary to encode and process information at the quantum level. Recent achievements in this direction have used either trapped ions^{1–3} or trapped photons in cavity quantum-electrodynamical systems^{3,4}. A third possibility that has been studied theoretically^{5,6} is to use trapped neutral atoms. Such schemes would benefit greatly from the ability to trap and address individual atoms with high spatial resolution. Here we demonstrate a method for loading and detecting individual atoms in an optical dipole trap of submicrometre size. Because of the extremely small trapping volume, only one atom can be loaded at a time, so that the statistics of the number of atoms in the trap, N , are strongly sub-poissonian ($\Delta N^2 \approx 0.5N$). We present a simple model for describing the observed behaviour, and we discuss the possibilities for trapping and addressing several atoms in separate traps, for applications in quantum information processing.

Our experiment is based on the well developed methods of neutral-atom trapping and cooling. These techniques have already been used to prepare large samples of atoms in specific motional quantum states⁷. Individual atoms have been stored for short times in high-finesse optical cavities^{8,9}, and for longer times in weakly confining optical traps^{10–12}. As far as quantum information

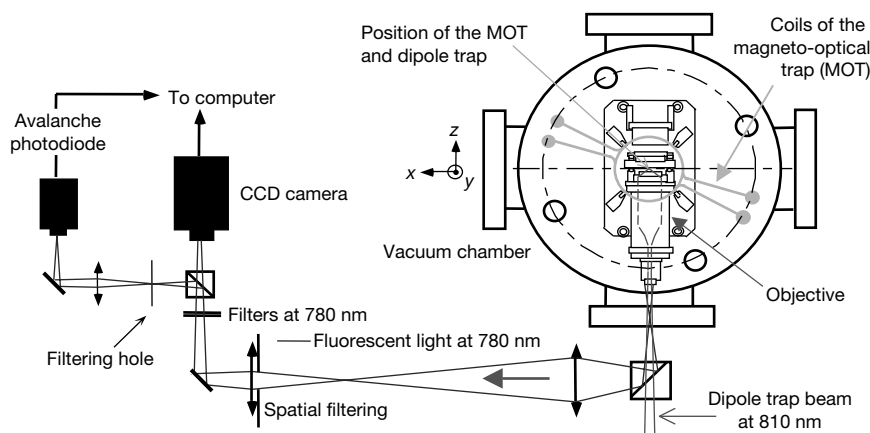


Figure 1 Main features of the experimental set-up. For simplicity, the slowing and MOT beams are not represented. A crucial feature of the experiment is the diffraction-limited (numerical aperture 0.7) focusing objective inside the vacuum chamber, which achieves submicrometre resolution. CCD, charge-coupled device.

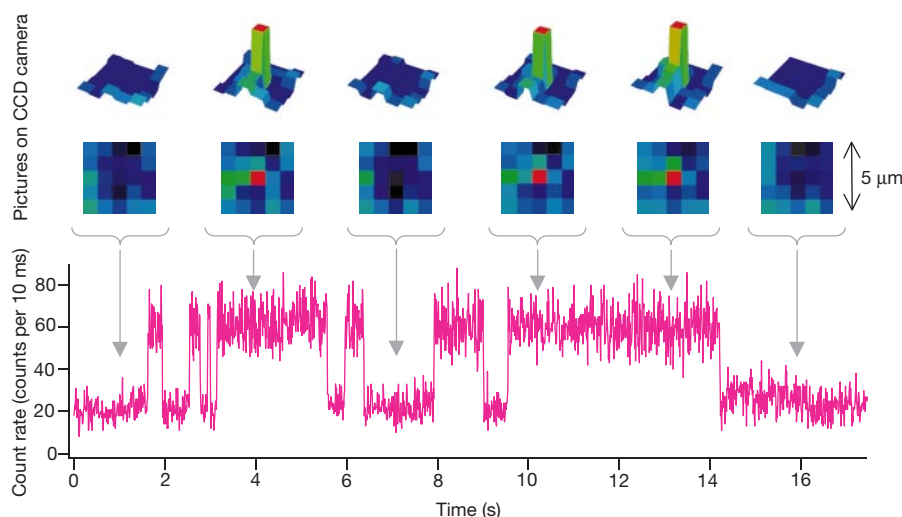


Figure 2 Single atom detection. Bottom, number of photons counted by an avalanche photodiode within 10-ms time bins. The steps correspond to either zero or one atom in the trap. Top, images of the trapped atom using a CCD camera. Squares (five pixels in side)

around the trap image are shown, in a two- and three-dimensional view of selected images. The counting time for each image is 200 ms. The appearance of one trapped atom can be seen in real time.

processing with atomic qubits is concerned, controlled cold collisions¹³ or atomic dipole–dipole coupling¹⁴ have been proposed as new mechanisms for pairwise entanglement of neutral atoms. Such schemes can in principle be implemented in optical lattices with controlled filling¹⁵, but loading and addressing individual trapping sites remains highly desirable^{5,6}. In order to demonstrate the trapping and detection of a single ⁸⁷Rb atom with submicrometre resolution, we use a red-detuned optical dipole trap, which attracts the atoms towards regions of high field intensity, corresponding to the focus of the beam. A specific feature of our trap—which has not, to our knowledge, been achieved before—is an extremely small focal spot, which has a beam waist radius of $w_0 < 1 \mu\text{m}$ (see Methods section). It is thus possible to reach very high intensities (100 kW cm^{-2}) with a moderate input power (3 mW), and to confine the atom within a very small volume.

When both the dipole trap and the pre-cooling magneto-optical trap (MOT, see Fig. 1 and Methods section) are turned on together, a bright spot appears on a single pixel of the imaging CCD (charge-coupled device) camera (Fig. 2). A more detailed fit to the data shows that the fluorescent spot has a full-width at half-maximum of $0.7 \mu\text{m}$ in the object plane, which is the diffraction limit of the imaging system. For moderate loading of the trap and a dipole trap power of 1 mW, each atom scatters about $4,000 \text{ photons s}^{-1}$

from the MOT beams. A blinking spot due to individual atoms entering and leaving the trap can thus easily be seen on the camera, using a counting time of 200 ms. The spot from the dipole trap is also imaged onto an avalanche photodiode (APD), which continuously counts photons within adjacent time bins. The detected fluorescence signal then exhibits characteristic steps, corresponding to individual atoms entering and leaving the trap. This behaviour is shown in Fig. 2, together with the CCD images, which are registered simultaneously.

A specific feature of our data, when compared to similar experiments using larger traps¹², is that count rates corresponding to more than one atom never occur. This statement can be made quantitative by using the histogram of counts shown in Fig. 3. The experimental data are compared with the histogram that can be expected from a compound Poisson law, which is fitted onto the peaks for $N = 0$ and $N = 1$. From these curves it is clear that the number of atoms N is strongly sub-poissonian (because $N = 2$ is missing), while the number of photons is approximately poissonian. The slight excess noise in the light (relative intensity of noise $< 10\%$) can be attributed mostly to technical noise in the MOT lasers.

The key feature of our data is the sub-poissonian character of the atom number^{16,17}. By integrating the two fluorescence peaks shown in Fig. 3, we find that the probabilities of trapping N atoms

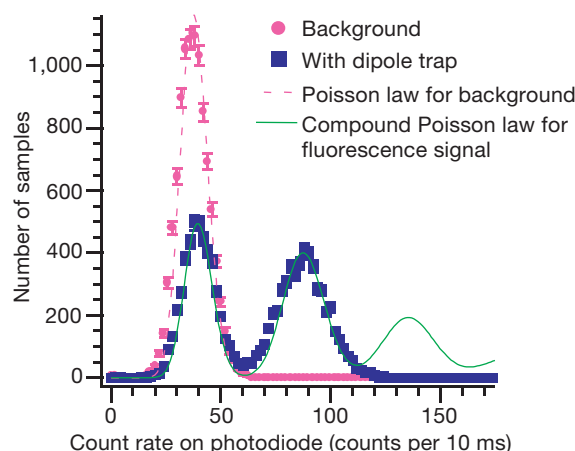


Figure 3 Sub-poissonian loading. The histogram of single-atom photon-counting data (Fig. 2) is displayed together with a compound Poisson law fitted to the 0-atom and 1-atom peaks. The 2-atom peak is clearly missing. The stray light background without the dipole trap is shown for comparison. Both histograms correspond to a total of 10,000 counting samples.

are $p(N=0) \approx p(N=1) \approx 0.5$, and $p(N \geq 2) \approx 0$. This corresponds to $\Delta N^2 \approx 0.5N$, instead of the $\Delta N^2 = N$ that is expected for a Poisson law. Most of the observed features are very well reproduced by assuming that atoms enter the trap at random times, but that if an atom is already in the trap, both atoms escape. This model predicts that the '0 atom' and '1 atom' events have equal weights of 50%. The mechanism responsible for the escape can reasonably be attributed to collisions assisted by the MOT light, which have been shown to be crucial in larger dipole traps^{18–20}. In practice, there might also be a collision with fast background atoms—this creates a trap lifetime which is typically a few seconds in our vacuum conditions.

A simple model combining this trap decay and the previously mentioned destructive two-body decay predicts a faster and faster swapping between no atoms and one atom in the trap when the loading rate is increased. We have checked this behaviour by measuring the autocorrelation of the detected fluorescence signal as a function of the loading rate (Fig. 4), and the observed results confirm that the correlation time is decreasing with increasing loading rate. In Fig. 4 we also show histograms—similar to those in Fig. 3—for various values of the magnetic field that controls the atomic density in the MOT (see Methods section). This confirms the following behaviour: for small loading rates, the lifetime of the trapped atom is equal to a constant value τ , essentially limited by collisions with fast atoms or molecules of the background gas, and the histograms show a majority of zero-atom periods, with long (several seconds) single-atom fluorescence periods. When the loading rate increases and becomes comparable to or larger than $1/\tau$, the probability of '1 atom' events increases, and quickly reaches its limit value of 50%. In this regime, the lifetime of the dipole trap decreases as the loading rate increases, following the statistics of the arrival of new atoms in the trap capture area.

For long-term storage of a single atom, the MOT is turned off 10 ms after an atom enters the trap, thereby stopping the loading and cooling processes, and only keeping the atom in the dipole trap. By turning on the MOT again after a given time, and measuring the probability that the atom is still in the dipole trap, the lifetime of the stored atom has been measured to be 2 s under typical experimental conditions.

Our set-up opens possibilities of exploring the proposed schemes for atom–atom entanglement. By simply sending another trapping beam at a small angle, and using the same optics, we have trapped two atoms at a controlled distance in the range 1–10 μm . After turning off the MOT beams, qubits can be implemented by using

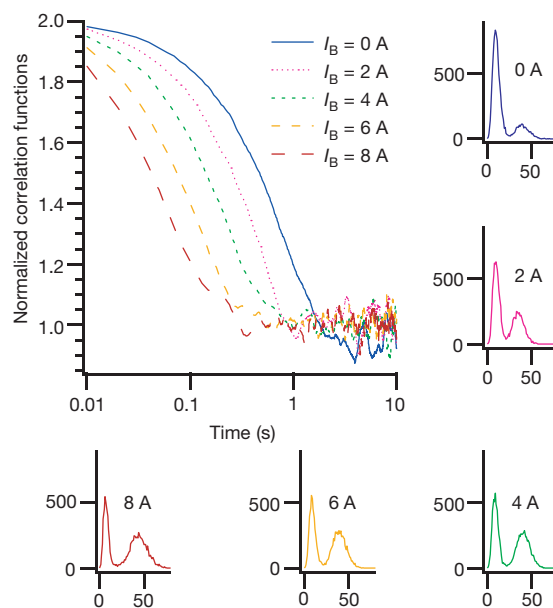


Figure 4 Intensity autocorrelations $g(t)$ of the MOT light emitted by the trapped atom (normalized for $g(0) = 2$ and $g(\infty) = 1$). The time behaviour is displayed together with selected histograms similar to the one shown in Fig. 3 (same axis labelling). The various curves are taken by increasing the MOT magnetic field B and thus the dipole trap loading rate. The current I_B for each magnetic field is given in amperes. The decreased storage time due to collisions leads to a faster swapping between zero and one atom. The histograms show again that there is never more than one atom in the trap.

hyperfine sublevels of the atomic ground state^{5,6,12}. Decoherence time, induced mostly by residual spontaneous emission from the dipole trap, and by intensity fluctuations of the trapping lasers, should be in the range 10–100 ms. One-qubit operations could be achieved by inducing Raman transitions between the hyperfine sublevels, separated by $\Delta\nu_H = 6.83$ GHz. Fast (microsecond) operation could be achieved by using as Raman beams the trapping laser itself, together with a pulsed control beam; the control beam would address each atom, and be detuned from the trapping beam by $\Delta\nu_H$.

For two-qubit operations, it should be possible to selectively couple one of the hyperfine sublevels to a high- n Rydberg state, by using a direct two-photon transition involving a pulsed blue laser and the trapping beam. Quite large two-photon effective Rabi frequencies would be easily accessible, owing to the very large intensity of the dipole trap beams. We have calculated that for two atoms separated by 3 μm , the dipole–dipole interaction between Rydberg states creates a level shift larger than 10 MHz. Moreover, a two-photon Rabi frequency of 200 kHz between the ground and Rydberg state should be accessible. Our set-up thus fulfils the basic requirements for realizing a fast C-NOT gate²¹, with a switching time in the microsecond range. In principle, a three-atom scheme (Toffoli gate) could also be implemented along the same lines. $\square$

Methods

Magneto-optical trap

The magneto-optical trap (MOT) uses a standard retroreflected 6 beam σ_+/σ_- configuration, except that the two beams within the plane of the coils (see Fig. 1) make a small angle of 20° (instead of the usual 90°), in order to fit within the optics of the focusing objective. The MOT is loaded from an atomic beam, slowed down by the usual chirped cooling method. Typically the MOT confines 10^7 atoms in a 1-mm-diameter cloud, with a density of about 10^{10} atoms cm^{-3} , or 0.01 atoms μm^{-3} . When correctly adjusted, the experiment can actually work without the slowing beams: the very small number of atoms that are captured by the MOT directly from the thermal atomic beam are enough to load the dipole trap. The loading rate is then approximately proportional to the MOT magnetic field; this is used for the experiment shown in Fig. 4.

Optics

The focusing optics consists of two parts, which first focus and then re-collimate the beam for diagnostic purposes (see Fig. 1). Both parts have a numerical aperture of 0.7, a working distance of 1 cm, and a common focal point set at the centre of the MOT. The focusing objective consists of nine lenses, and achieves a diffraction-limited resolution of 0.7 μm , while the re-collimating objective consists of four lenses. The fluorescence light induced by the MOT beams is collected by the focusing objective, and gives a magnified image of the trap on a CCD camera. As the CCD camera has a slow response time, an APD is used in parallel, and monitors only the light coming from the trap region. Typically, the 1- μm -diameter dipole trap is imaged onto a 50- μm pinhole with a magnification of 20. The pulses from the APD are continuously registered by a computer, and can be processed to get histograms, intensity correlations, or photon counting statistics within the counting window.

Dipole trap

The linearly polarized dipole trap beam from a frequency stabilized titanium-sapphire laser is brought to the set-up using an optical fibre. Its wavelength is 810 nm, to be compared with the 795 nm (D1) and 780 nm (D2) lines of rubidium, so the trap operates far off resonance²², where the spontaneous emission is very small. The typical oscillation frequency in the trap along the x - y (transverse) axes is calculated to be close to 200 kHz; this value has been confirmed experimentally by using parametric excitation of the oscillatory motion. The MOT cooling is expected to be working over a significant fraction of the potential well of the dipole trap. The trapped-atom temperature can be evaluated by measuring the fluorescence decay when turning off the MOT and dipole trap beams, and after some time turning on the MOT again. Data from the present experimental set-up, which is being improved, show that the trapped atom temperature is higher than 50 μK . An upper limit is given by the smallest potential for reliably trapping one atom, which is at present 1 mK. More accurate measurements of the temperature and motional properties of the trapped atom are under way.

Received 14 March; accepted 11 May 2001.

1. Sackett, C. A. *et al.* Experimental entanglement of four particles. *Nature* **404**, 256–259 (2000).
2. Steane, A. *et al.* Speed of ion-trap quantum-information processors. *Phys. Rev. A* **62**, 042305–1–042305–9 (2000).
3. Walther, H. Single atom experiments in cavities and trap. *Proc. R. Soc. Lond. A* **454**, 431–445 (1998).
4. Rauschenbeutel, A. *et al.* Coherent operation of a tunable quantum phase gate in cavity QED. *Phys. Rev. Lett.* **83**, 5166–5169 (1999).
5. Calarco, T., Briegel, H. J., Jaksch, D., Cirac, J. I. & Zoller, P. Entangling neutral atoms for quantum information processing. *J. Mod. Opt.* **47**, 2137–2149 (2000).
6. Brennen, G. K., Deutsch, I. H. & Jessen, P. S. Entangling dipole-dipole interactions for quantum logic with neutral atoms. *Phys. Rev. A* **61**, 062309–1–062309–10 (2000).
7. Moringa, M., Bouchoule, I., Karam, J. C. & Salomon, C. Manipulation of motional quantum states of neutral atoms. *Phys. Rev. Lett.* **83**, 4037–4040 (1999).
8. Doherty, A. C., Lynn, T. W., Hood, C. J. & Kimble, H. J. Trapping of single atoms with single photons in cavity QED. *Phys. Rev. A* **63**, 013401–1–013401–24 (2001).
9. Pinkse, P. W. H., Fischer, T., Maunz, P. & Rempe, G. Trapping an atom with single photons. *Nature* **404**, 365–368 (2000).
10. Hu, Z. & Kimble, H. J. Observation of a single atom in a magneto-optical trap. *Opt. Lett.* **19**, 1888–1890 (1994).
11. Ruschewitz, F., Bettermann, D., Peng, J. L. & Ertmer, W. Statistical investigations on single trapped neutral atoms. *Europhys. Lett.* **34**, 651–656 (1996).
12. Frese, D. *et al.* Single atoms in an optical dipole trap: towards a deterministic source of cold atoms. *Phys. Rev. Lett.* **85**, 3777–3780 (2000).
13. Calarco, T. *et al.* Quantum gates with neutral atoms: Controlling collisional interactions in time-dependent traps. *Phys. Rev. A* **61**, 022304–1–022304–11 (2000).
14. Brennen, G. K., Caves, C. M., Jessen, P. S. & Deutsch, I. H. Quantum logic gates in optical lattices. *Phys. Rev. Lett.* **82**, 1060–1063 (1999).
15. DePue, M. T. *et al.* Unity occupation of sites in a 3D optical lattice. *Phys. Rev. Lett.* **82**, 2262–2265 (1999).
16. Rempe, G. & Walther, H. Sub-Poissonian atomic statistics in a micromaser. *Phys. Rev. A* **42**, 1650–1655 (1990).
17. Orzel, C., Tuchman, A. K., Fenselau, M. L., Yasuda, M. & Kasevich, M. A. Squeezed states in a Bose-Einstein condensate. *Science* **291**, 2386–2389 (2001).
18. Gensemer, S. D., Gould, P. L., Leo, P. J., Tiesinga, E. & Williams, C. J. Ultracold ^{87}Rb ground-state hyperfine-changing collisions in the presence and absence of laser light. *Phys. Rev. A* **62**, 030702–1–030702–4 (2000).
19. Nesidal, R. C. & Walker, T. G. Light-induced ultracold spin-exchange collisions. *Phys. Rev. A* **62**, 030701–1–030701–4 (2000).
20. Kuppens, S. J. M., Corwin, K. L., Miller, K. W., Chupp, T. E. & Wieman, C. E. Loading an optical dipole trap. *Phys. Rev. A* **62**, 013406–1–013406–13 (2000).
21. Jaksch, D. *et al.* Fast quantum gates for neutral atoms. *Phys. Rev. Lett.* **85**, 2208–2211 (2000).
22. Miller, J. D., Cline, R. A. & Heinzen, D. J. Far-off-resonance optical trapping of atoms. *Phys. Rev. A* **47**, R4567–R4570 (1993).

Acknowledgements

The contributions of K. Vigneron, H. Wilhelm and T. Zhang to early stages of the experiment are acknowledged. This work was supported by the European IST/FET programme 'QUBITS' and by the European IHP network 'QUEST'.

Correspondence and requests for materials should be addressed to P.G.
(e-mail: philippe.grangier@iota.u-psud.fr).

Nonlinear limits to the information capacity of optical fibre communications

Partha P. Mitra & Jason B. Stark

Bell Laboratories, Lucent Technologies, Murray Hill, New Jersey 07974, USA

The exponential growth in the rate at which information can be communicated through an optical fibre is a key element in the 'information revolution'. However, as for all exponential growth laws, physical limits must be considered. The nonlinear nature of the propagation of light in optical fibre has made these limits difficult to elucidate. Here we use a key simplification to investigate the theoretical limits to the information capacity of an optical fibre arising from these nonlinearities. The success of our approach lies in relating the nonlinear channel to a linear channel with multiplicative noise, for which we are able to obtain analytical results. In fundamental distinction to linear channels with additive noise, the capacity of a nonlinear channel does not grow indefinitely with increasing signal power, but has a maximal value. The ideas presented here may have broader implications for other nonlinear information channels, such as those involved in sensory transduction in neurobiology. These have been often examined using additive noise linear channel models¹ but, as we show here, nonlinearities can change the picture qualitatively.

The classical theory of communications¹ was developed mostly in the context of linear channels with additive noise, which was adequate for electromagnetic propagation through wires and cables that have until recently been the main conduits for information flow. Fading channels or channels with multiplicative noise have been considered, for example in the context of wireless communications², although such channels remain theoretically less tractable than the additive noise channels. However, with the advent of optical fibre communications we are faced with a nonlinear propagation channel that poses major challenges to our understanding. The difficulty is that the input-output relationship of an optical fibre channel is obtained by integrating a nonlinear partial differential equation and may not be represented by an instantaneous nonlinearity. Channels where the nonlinearities in the input-output relationship are not instantaneous are in general ill understood, the optical fibre simply being a case of current relevance. The understanding of such nonlinear channels with memory are of fundamental interest, both because communication rates through optical fibre are increasing exponentially and we need to know where the limits are, and also because understanding such channels may give us insight elsewhere, such as into the design principles of neurobiological information channels at the sensory periphery.

The capacity of a communication channel is the maximal rate at which information may be transferred through the channel without error. The capacity can be written as a product of two conceptually distinct quantities, the spectral bandwidth W and the maximal spectral efficiency which we will denote C' . In the classic capacity formula for the additive white gaussian noise channel with an average power constraint, $C = W \log_2(1 + S/N)$, the spectral bandwidth W , which has dimensions of inverse time, multiplies the dimensionless maximal spectral efficiency $C' = \log_2(1 + S/N)$ (sometimes written with 'units' of bits $\text{s}^{-1} \text{Hz}^{-1}$). Here S and N are the signal and noise powers respectively. As the maximal spectral efficiency is logarithmic in the signal to noise ratio (S/N), the capacity is principally determined by the bandwidth W . In the case of an optical fibre, the intrinsic loss mechanisms of light propagating

through silica fundamentally limits W to a maximum of about 50 THz (ref. 3) corresponding to a wavelength range of about 400 nm (1.2–1.6 μm). This is to be compared with current systems where the total bandwidth is limited to about 15 THz. If the channel was linear, the maximal spectral efficiency would be $C' = \log_2(1 + S/N)$, S being input light intensity and N the intensity of amplified spontaneous emission noise in the system. An output S/N of say 100 (that is, 20 dB), would then yield a spectral efficiency of 6.6, which for a 50 THz channel would correspond to a capacity of 330 Tbit s^{-1} . The channel, of course, is not linear; so how do the nonlinearities affect the spectral efficiency of the fibre? Our basic conclusion here is that the impact is severe and qualitative. As shown in Fig. 1, the effect is a saturation and eventual decline of spectral efficiency as a function of input signal power, in complete contrast with the linear channel case. We now proceed to discuss this result.

We concentrate on the most important propagation nonlinearity, namely the intensity dependence of the refractive index⁴ n : $n = n_0 + n_2 I$. Here n_0 is the linear refractive index and n_2 is a constant. The nonlinearity is weak, but its effects accumulate over long propagation distances. Three principal parameters are of interest: the group velocity dispersion $\beta \approx 10 \text{ ps}^2 \text{ km}^{-1}$, the propagation loss $\alpha \approx 0.2 \text{ dB km}^{-1}$ and the strength of the nonlinear refractive index, $\gamma \approx 1 \text{ W}^{-1} \text{ km}^{-1}$. The propagation loss is compensated by interposing optical amplifiers into the system, which inject spontaneous emission noise into the system with strength⁵ $I_1 = aGh\nu B$. Here G is the amplifier gain, h is Planck's constant, ν and B are the centre frequency and bandwidth respectively, and a is a numerical constant (which we assume to be 2). For n_s spans of fibre with interspersed amplifiers, absorption may be accounted for by considering an effective length, $L_{\text{eff}} \approx n_s/\alpha$. In the absence of nonlinearity, the maximal spectral efficiency is $C'_0 = \log_2(1 + I/I_n)$, I being the input power and $I_n = n_s I_1$ being the total additive noise power. We note that C'_0 vanishes as the inverse power of the system length.

For a variety of reasons, principally limitations in the electronic bandwidth, it is impractical to modulate the full optical bandwidth at once. Current attempts towards achieving maximal information throughput involve wavelength division multiplexing (WDM), where the whole optical bandwidth is broken up into disjoint

frequency bands each of which is modulated separately. We confine our attention to such systems (which from an information theory perspective corresponds to the 'multi-user' case)⁶, although we also comment on the ideal case of a single data stream (the 'single-user' case).

The nonlinear propagation effects in the evolution of the electric field amplitude involve a cubic term in the electric field. In a WDM system, the nonlinearities are classified by the field amplitudes participating in this cubic term. For self-phase modulation (SPM), all three fields belong to the channel of interest; for cross-phase modulation (CPM), two fields belong to a different channel; and for four-wave mixing (FWM), all three amplitudes belong to other channels. Of these, FWM gives rise to additive noise to the channel of interest but is strongly suppressed by dispersion and will be omitted from the present considerations for simplicity. Its effects can be incorporated by including an additive term in the model below, and we comment on this at a later stage. We also neglect SPM, as these effects are deterministic for the given channel and could be reduced using nonlinear precompensation. Finally, we are left with CPM, which appears to be the principal source of nonlinear capacity impairment in the multi-user case for realistic parameter ranges. It causes multiplicative noise, which gives rise to qualitatively new effects and severe impairments in the channel capacity compared to additive noise. When channels are so close as to have almost overlapping frequency components, the sharp distinction between CPM and FWM is lost, owing to non-phase-matched terms that cannot be simply categorized. However, the distinction between multiplicative and additive noise can still be made, and here we group all multiplicative noise components (that is, those involving an amplitude from the channel of interest) under CPM.

We note that although CPM arises from a deterministic propagation equation of the full electric field, owing to the multi-user nature of WDM it nevertheless impairs the capacity; this effect is analogous to multi-user interference in wireless communication systems, except that in the present case, propagation nonlinearities cause users to interfere even though they have non-overlapping frequency bands.

We model the propagation channel in the presence of CPM by means of a linear Schrödinger equation with a random potential

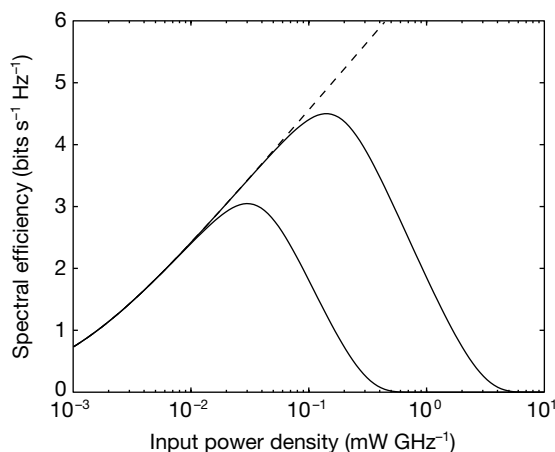


Figure 1 Spectral efficiencies. The curves represent lower bounds to the spectral efficiency for a homogenous length of fibre for a multi-user WDM system, given analytically by equation (2). Although the curves represent lower bounds, we argue in the text that the true capacity shows the same qualitative non-monotonic behaviour with respect to input signal powers. The spectral efficiencies displayed in the figure correspond to the capacity per unit bandwidth, $C' = C/(\eta_c \delta \nu)$. Here $\delta \nu$ includes both the channel bandwidths and the inter-channel spacing. The parameters used for the figure are

$n_c = 100$, $L_{\text{eff}} = 100 \text{ km}$, $D = 20 \text{ ps nm}^{-1} \text{ km}^{-1}$, $\delta \nu = 1.5B$ where $B = 10 \text{ GHz}$ is the individual channel width. The two continuous curves correspond to $\gamma = 1 \text{ W}^{-1} \text{ km}^{-1}$ and $\gamma = 0.1 \text{ W}^{-1} \text{ km}^{-1}$. The lower curve corresponding to $\gamma = 1$. The spontaneous noise strength I_1 is computed for the formula $I_1 = aGh\nu B n_s$ as explained in the text, with $a = 2$, $G = 1,000$, $\nu = 200 \text{ THz}$. The dashed curve represents the spectral efficiencies of the corresponding linear channels given by $\gamma = 0$.

fluctuating both in space and time. This is easily justified, starting from the nonlinear Schrödinger equation description commonly used to describe light propagation in single-mode optical fibres⁴. If only CPM effects are retained, the propagation equation for the field amplitude E_i in channel i is given by:

$$i\partial_z E_i = \frac{\beta}{2} \partial_t^2 E_i + V(z, t) E_i \quad (1)$$

where $V(z, t) = -2\gamma \sum_{j \neq i} |E_j(z, t)|^2$. Here β is the dispersion coefficient, t is time, z measures distance along the fibre, and ∂_z and ∂_t are partial derivatives with respect to z and t . Because independent streams of information are transmitted in the channels $j \neq i$, $V(z, t)$ appears as a random noise term to the channel i . We note that the nonlinear propagation equation has now been reduced to a linear Schrödinger equation with a stochastic potential, so that the nonlinear channel has become a channel with multiplicative noise. We now need an adequate model for the stochastic properties of $V(z, t)$. If the dispersion is substantial, we propose that $V(z, t)$ may be approximated by a gaussian stochastic process that is short-range-correlated in both space and time. As V is obtained by adding a large number of different channels, each of which is short-range-correlated in time ($\tau \approx 1/B$, where B is the channel bandwidth), we can expect V to have a correlation time of approximately $1/B$. Dispersion causes the channels to travel at different speeds, thus causing V to be short-range-correlated in space as well, with a correlation length related to the dispersion length. Because V is a sum of intensities, it has a non-zero mean, so we define $\delta V(z, t) = V(z, t) - \langle V \rangle$, where $\langle V \rangle$ denotes the average value of V . Removing a constant for the potential causes an overall phase shift independent of space and time, which is irrelevant to the present considerations.

The parameter of interest in the following is the integrated strength of the fluctuating field, $\eta = \int dz \langle \delta V(z, 0) \delta V(0, 0) \rangle$. In order to estimate η , we consider a simplified propagation model for the channels other than the one of interest, in which nonlinearities are neglected, and stochastic bit streams at the inputs to the channels are propagated forward with constant group velocities. The group velocity difference between two channels separated by a spacing $\Delta\lambda$ is $D\Delta\lambda$. In this model with n_c other channels evenly spaced by $\Delta\lambda$ around the channel of interest, each with intensity I and bandwidth B , we obtain $\eta = 2\ln(n_c/2)(\gamma I)^2/(BD\Delta\lambda)$. Here D is the dispersion parameter $D = -2\pi c\beta/\lambda^2$. Although this is a simplified model for the other channels, numerical simulations of propagation, including the nonlinearities and dispersion for the side channels, show that the estimate of η is accurate.

We note that the denominator in the expression of η is the inverse of the dispersion length L_D for the given channel spacing. This form for η follows from assuming that $L_{\text{eff}} \gg L_D$, because in this limit the integral defining η is cut off by L_D . If, on the other hand, $L_{\text{eff}} \leq L_D$, the integral would be cut off by L_{eff} , so that one would have to replace L_D by L_{eff} in the equation for η . Dispersion causes channels at spacing $\Delta\lambda_j$ to be suppressed proportionately to $1/(\Delta\lambda_j) \propto 1/j$, which leads to the logarithmic factor after summation.

The electric field evolution is given in terms of a propagator $U(t, t'; L)$, obtained by integrating the stochastic Schrödinger equation. We model the amplifier noise as an additive term with strength I_n as defined earlier. The channel is specified in terms of a relation between the input and output electric field amplitudes, $E_{\text{out}}(t) = \int dt' U(t, t'; L) E_{\text{in}}(t') + n(t)$. U is stochastic, so, owing to the underlying stochasticity of $V(z, t)$, the model corresponds to a channel with multiplicative noise. An exact capacity computation is still intractable, but an analytic lower bound may now be obtained, based on the following information theoretic result (E. Telatar, personal communications): the capacity C' of a channel with input X and output Y related by a conditional distribution $p(Y|X)$ and an input power constraint $E(|X|^2) = P$ satisfies the inequalities

$C' = \max_{p(X)} I(X, Y) \geq I(X_G, Y) \geq I(X_G, Y_G)$. Here $I(X_G, Y)$ is the mutual information when $p(X)$ is chosen to be $p_G(X)$, a gaussian satisfying the power constraint; $I(X_G, Y_G)$ is the mutual information of a pair (X_G, Y_G) with the same covariance matrix as the pair (X_G, Y) . The first inequality is trivial, as $p_G(X)$ is not necessarily the optimal input distribution. A proof of the second inequality is outlined in the Methods section.

The quantity $I(X_G, Y_G)$ may be computed from the correlators $\langle E_{\text{in}}(t) E_{\text{in}}^*(t') \rangle$, $\langle E_{\text{out}}(t) E_{\text{out}}^*(t') \rangle$ and $\langle E_{\text{out}}(t) E_{\text{in}}^*(t') \rangle$. The first is defined *a priori* through the assumption of band-limited gaussian white noise input with a power constraint. This is important for the application of the result above which requires a gaussian distribution for X . The second follows from the first using the unitarity of U . The third correlator requires computation of the average propagator $\langle U \rangle$, where the average is over realizations of $V(z, t)$. For a gaussian, delta-correlated V , we obtain $\langle U(t, t'; L) \rangle = \exp(-\eta L/2) U_0(t - t'; L)$ (see Methods), where U_0 is the propagator for $V = 0$. Assembling these results, we finally obtain an analytic expression for a lower-bound C_{LB} to the channel capacity of the stochastic Schrödinger equation model:

$$C_{\text{LB}} = n_c B \log_2 \left(1 + \frac{e^{-(I_0/2)} I}{I_n + (1 - e^{-(I_0/2)}) I} \right) \quad (2)$$

where I_0 is given by

$$I_0 = \sqrt{\frac{BD\Delta\lambda}{2\gamma^2 \ln(n_c/2) L_{\text{eff}}}} \quad (3)$$

The fundamental departure from a linear channel in the above capacity expression is the appearance of a nonlinear intensity scale I_0 . For $B = 40$ GHz, $D = 20$ ps nm⁻¹ km⁻¹, $\Delta\lambda = 1$ nm, $\gamma = 1$ W⁻¹ km⁻¹, $n_c = 100$, and $L_{\text{eff}} \approx n_c/\alpha = 100$ km, we get $I_0 = 32$ mW. Examination of equation (3) shows that I_0 has reasonable dependence on all relevant parameters: namely, it increases with dispersion, the bandwidth and the channel spacing, but decreases with increasing system length and number of channels.

The reason for the non-monotonic behaviour of the capacity estimate is that if we consider any particular channel, the signal in the other channels appear as noise in the channel of interest, owing to the nonlinearities. This 'noise' power increases with the 'signal' strength, thus causing degradation of the capacity at large 'signal' strength. The behaviour of equation (2) is graphically illustrated in Fig. 1, where the spectral efficiency (bits transmitted per second per unit bandwidth) is shown as a function of input power. The peak capacity, achieved for an intensity $I_{\text{max}} \approx (I_0^2 I_n/2)^{1/3}$, is given approximately by $C_{\text{max}} \approx \frac{2}{3} n_c B \log_2(2I_0/I_n)$.

We note that at large intensities, multiplicative noise causes an exponential attenuation of the effective signal power, which is why it dominates over nonlinear additive noise: the latter would contribute an additive term to the denominator in equation (2) that would be cubic in intensity. At large intensities where this term is important, the exponential in the numerator arising from multiplicative noise dominates.

If the input intensity is kept fixed, the capacity bound declines exponentially with the system length, in contrast with the power-law decay for the linear propagation channel. This is as expected, because correlations of the electric field decay exponentially due to the fluctuating potential in equation (1). On the other hand, the maximal spectral efficiency given by C_{max} declines only logarithmically.

Finally, we present qualitative arguments as to why the single-user case is expected to show the same non-monotonicity of spectral efficiency with the input signal intensities. In the multi-user case, unknown information in other channels leads to multiplicative noise. In the single-user case, the cubic nonlinearity is a deterministic process that would not by itself degrade channel capacity

except to a small extent through loss of power out of band. However, noise photons injected by the amplifiers and random photon deletion due to absorption cause the propagating field to acquire a stochastic component, which also leads to multiplicative noise effects through the nonlinear term. This leads us to believe that the single-user case will qualitatively exhibit the same decline in spectral efficiency with intensity as the multi-user case. The same qualitative arguments apply to interposing optical phase conjugation elements for nonlinear compensation; fibre absorption and amplifier noise can still be shown to cause multiplicative noise through the cubic term. It would be incorrect to conclude that nonlinearities must always impair capacity: signal regenerators are an obvious counterexample. However, the fundamental insight in the current work is that in nonlinear propagation channels, qualitatively new phenomena that arise from multiplicative noise effects can severely degrade capacity. $\square$

Methods

Gaussian bound to the channel capacity

Proof of the inequality $I(X_G, Y) \geq I(X_G, Y_G)$: define $p(X, Y)$ as the product $p_G(X)p(Y|X)$, and $p_G(X, Y)$ to be the joint gaussian distribution having the same second moments as $p(X, Y)$. Also define $p_G(Y)$ to be the corresponding marginal of $p_G(X, Y)$.

$$I(X_G, Y) = \int dXdY p(X, Y) \log \left(\frac{p(X, Y)}{p_G(X)p(Y)} \right) \quad (4)$$

$$= \int dXdY p(X, Y) \left[\log \left(\frac{p_G(X, Y)}{p_G(X)p_G(Y)} \right) - \log \left(\frac{p_G(X, Y)p_G(Y)}{p(X, Y)p_G(X)} \right) \right] \quad (5)$$

As $p(X, Y)$ and $p_G(X, Y)$ share second moments, the first term on the right-hand side is $I(X_G, Y_G)$. The second term may be simplified using the convexity of the logarithm, $\langle \log(f) \rangle \leq \log(\langle f \rangle)$ to obtain

$$I(X_G, Y) \geq I(X_G, Y_G) - \log \left[\int dXdY p_G(X, Y) \frac{p(Y)}{p_G(Y)} \right] \quad (6)$$

$$\geq I(X_G, Y_G). \quad (7)$$

The second inequality follows by first performing the integral over X , and noting that $\log(\int dY p(Y)) = \log(1) = 0$.

Derivation of the average propagator $\langle U \rangle$

This can be done by resumming the perturbation series exactly for $\langle U \rangle$, for delta-correlated $V(z, t)$. Alternatively, using path integrals⁷, $\langle U(t, t'; L) \rangle = U_0(t - t'; L) \langle \exp(i \int_0^L dz V(z, t(z))) \rangle$, where the average is taken over V as well as over paths $t(z)$ satisfying $t(0) = t$, $t(L) = t'$. The result in the text follows by performing the gaussian average over V . Because $\phi = \int_0^L dz V(z, t(z))$ is a linear combination of gaussian variables, it is also gaussian-distributed and satisfies $\langle \exp(i\phi) \rangle = \exp(-\langle \phi^2 \rangle / 2)$. The result follows by noting that for delta-correlated V , $\langle \phi^2 \rangle$ is a constant given by ηL . The delta correlations need to be treated carefully; this can be done by smearing the delta functions slightly and leads to the definition of η given earlier in the text.

Received 13 November 2000; accepted 19 April 2001.

- Shannon, C. E. A mathematical theory of communications. *Bell Syst. Tech. J.* **27**, 379–423, 623–656 (1948).
- Biglieri, E., Proakis, J. & Shamai, S. Fading channels: Information-theoretic and communications aspects. *Inform. Theory Trans.* **44**, 2619–2692 (1998).
- Kogelnik, H. High-capacity optical communication. *IEEE Sel. Topics Quant. Elec.* **6**(6), 1279 (2000).
- Agrawal, G. P. *Nonlinear Fiber Optics* (Academic, San Diego, 1995).
- Desurvire, E. *Erbium Doped Fibre Amplifiers* 69 (Wiley, New York, 1994).
- Cover, T. M. & Thomas, J. A. *Information Theory* 374–458 (Wiley, New York, 1991).
- Feynman, R. P. & Hibbs, R. A. *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1963).

Acknowledgements

We gratefully acknowledge extensive discussions with E. Telatar, and also with A. Green, P. B. Littlewood, R. Slusher, A. Chraplyvy and G. Foschini. We thank M. Povinelli and L. Wegener for performing numerical and analytical computations to verify the considerations in this Letter, and D. R. Hamann and R. Slusher for careful readings of the manuscript.

Correspondence and requests for materials should be addressed to P.P.M. (e-mail: pmitra@bell-labs.com).

Control of conformational and interpolymer effects in conjugated polymers

J. Kim^{*†} & T. M. Swager^{‡†}

^{*} Department of Materials Science and Engineering, [†] Center for Materials Science and Engineering, and [‡] Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

The role of conjugated polymers in emerging electronic, sensor and display technologies is rapidly expanding. In spite of extensive investigations^{1–11}, the intrinsic spectroscopic properties of conjugated polymers in precise conformational and spatial arrangements have remained elusive. The difficulties of obtaining such information are endemic to polymers, which often resist assembly into single crystals or organized structures owing to entropic and polydispersity considerations. Here we show that the conformation of individual polymers and interpolymer interactions in conjugated polymers can be controlled through the use of designed surfactant poly(*p*-phenylene-ethynylene) Langmuir films. We show that by mechanically inducing reversible conformational changes of these Langmuir monolayers, we can obtain the precise interrelationship of the intrinsic optical properties of a conjugated polymer and a single chain's conformation and/or interpolymer interactions. This method for controlling the structure of conjugated polymers and establishing their intrinsic spectroscopic properties should permit a more comprehensive understanding of fluorescent conjugated materials.

We designed and synthesized (see Supplementary Information for details) four poly(*p*-phenylene-ethynylene) compounds (PPEs) using four surfactant 'building blocks' that display preferential orientations at the air–water interface. Unique combinations of these building blocks (A–D in Fig. 1) allow us to control an isolated polymer chain's conformation and interpolymer interactions. Building block A, with two *para*-hydrophobic dioctylamide groups, is expected to display a face-on structure, with its phenyl groups co-facial to the air–water interface^{12,13}. B has two *para*-hydrophilic triethyleneoxide groups, and is also expected to prefer a face-on structure. The third building block, C, has one hydrophobic and one hydrophilic group *para* to each other, and is inclined towards an equilibrium edge-on structure with the π -plane normal to the interface. The last block, D, with two *ortho*-hydrophobic dodecyloxy groups, favours an edge-on structure. By design we have used these four building blocks to produce PPEs with precise structural features at the air–water interface. As we will show below, these polymers display one of three equilibrium organizations—face-on, alternating face-on and edge-on (zipper), or edge-on—depending upon the structure and surface pressure (Fig. 1).

Central to our analysis is the ability to switch between different structures, by applying mechanical force while monitoring the Langmuir monolayer's optical spectra. Figure 1 shows the chemical structures, conformations and spatial arrangements at the air–water interface of the PPEs. Our assignments of the conformations and spatial arrangements are based on a self-consistent analysis of the pressure–area (P–A) isotherms (Fig. 2a), *in situ* ultraviolet–visible (UV–vis.) and fluorescence spectroscopy (Fig. 3), and molecular modelling (Fig. 2b). Polymer 1 favours the face-on structure with a large extrapolated area per repeating unit of 240 Å². As the monolayer is compressed, the surface pressure increases until the monolayer folds into multilayers at 30 mN m^{–1} with an area per repeating unit of 130 Å². Molecular modelling predicts a repeating unit area of 126 Å² when 1 is highly compressed in the face-on structure (Fig. 2b). The wavelength of maximum

absorption (λ_{max}) of **1** at the air–water interface at 0 mN m^{-1} is red-shifted by 35 nm relative to its solution value owing to the increased π -conjugation length in the face-on structure^{5,14}. Our studies point to the polymer having a highly non-planar structure in solution (see below). Compression perturbs the π -conjugation system, causing the blue shift in absorption spectra with increased pressure (Fig. 3a). During the compression, the fluorescence quantum yield is unchanged, indicating—as expected—that edge-to-edge interpolymer interactions do not affect the quantum yield (Fig. 3b). After the monolayer of **1** folds into a multilayer at 30 mN m^{-1} , aggregated fibrils are visually observed. *In situ* UV–vis. spectroscopy also shows an additional aggregation peak at 464 nm, 13 nm to the red of the absorption λ_{max} of the face-on structure with the maximized π -conjugation. We confirmed that the 464-nm band is the result of interpolymer co-facial π – π interactions by conducting the same *in situ* UV–vis. experiment on a previously studied polymer¹⁵ that is isostructural except that the ethylene oxide side chains of B are part of a macrocycle that prevents strong interchain electronic interactions. In this case, even after the folding into a multilayer, the UV–vis. spectrum does not show the aggregation peak at 464 nm. We note that the polymer conformation in the aggregated state probably exhibits an extended conjugation, which will facilitate the strong interpolymer electronic interactions. The P–A isotherm of **1** is completely reversible, as are the spectroscopic changes.

Polymer **2** also shows completely reversible P–A isotherms and spectroscopic changes during cycles. The extrapolated area per repeating unit of $240 \text{ \AA}^2$ for **2** is the same as for **1**, confirming the face-on structure. A distinct difference of the P–A isotherm of **2** is

the transition between 15 and 20 mN m^{-1} , when the orientation of the C repeating units change to an edge-on structure, allowing further compression up to $90 \text{ \AA}^2$. We refer to **2**'s new alternating face-on and edge-on conformation (with neighbouring chains interlocked) as a zipper structure (Fig. 1). The areas per repeating unit predicted by molecular modelling are $196 \text{ \AA}^2$ and $134 \text{ \AA}^2$ for the face-on and the zipper structure respectively, which are in good agreement with the P–A isotherm (Fig. 2). The transition point ($140 \text{ \AA}^2$) in the P–A isotherm was calculated from the second derivative of the isotherm.

The spectroscopic measurements also support this zipper structure. Initial compression up to 15 mN m^{-1} causes the same minor perturbation of the π -conjugation system as in the case of **1**. However, continued compression from 15 to 20 mN m^{-1} causes a transition from the face-on to the zipper structure, which decreases the π -conjugation length thereby generating a large (21-nm) blue shift. A 34-nm blue shift is observed over the entire compression (0 – 20 mN m^{-1}). The fluorescence spectrum also showed a 9-nm blue shift with the face-on to zipper transition. The 38% increase in quantum yield is probably due to a larger energy gap and reduced intramolecular dynamics in the compressed state. The larger energy separation of the lowest singlet excited state, S_1 , and the ground state S_0 , decreases the Franck–Condon overlaps of the nuclear wavefunctions that control the internal conversion rate, and therefore an improved quantum yield is expected¹⁶. It is also possible that an interlocked zipper structure, which lacks extended π -conjugation, reduces the exciton's diffusion length thereby reducing quenching with molecular oxygen or other impurities. The λ_{max} of the initial face-on structure is red-shifted by 38 nm relative to solution,

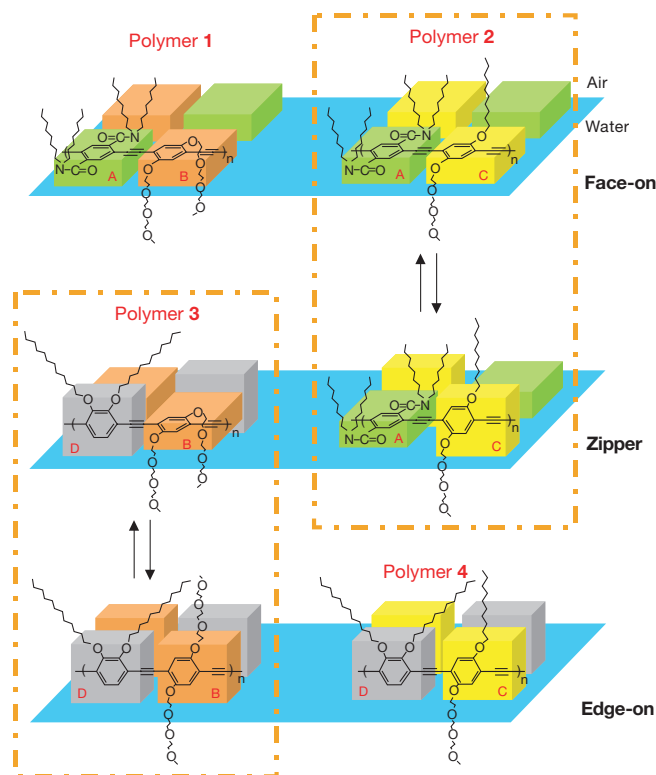


Figure 1 Conformations and spatial arrangements of polymers **1–4** at the air–water interface and their reversible conversions between face-on, zipper and edge-on structures. Polymer **1**: number-average molecular mass $M_n = 23,000$, Poly-dispersity index, PDI = 2.4. Polymer **2**: $M_n = 293,000$, PDI = 1.7. Polymer **3**: $M_n = 115,000$, PDI = 2.2. Polymer **4**: $M_n = 96,000$, PDI = 2.8. The monomers A, B, C and D are represented by green, orange, yellow and grey boxes, respectively.

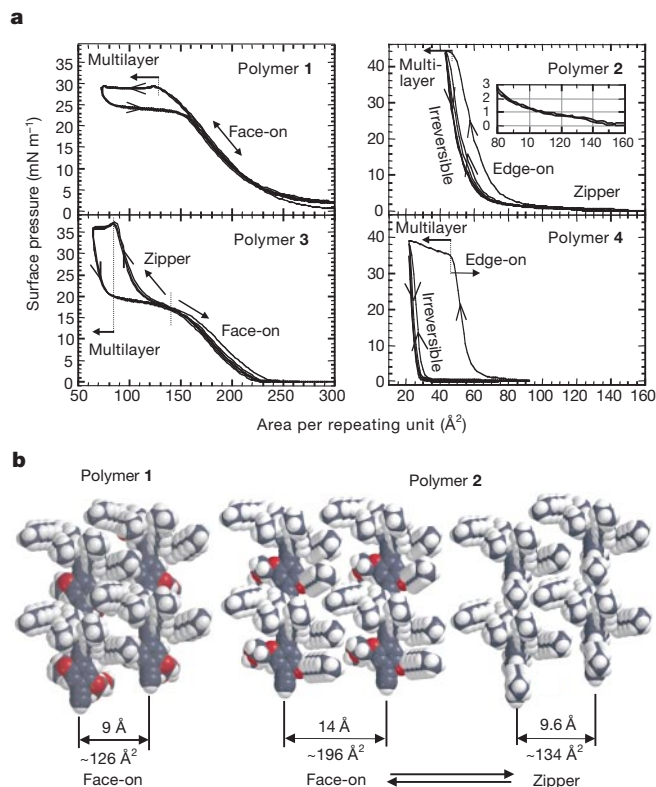


Figure 2 Pressure–area (P–A) isotherms of polymers **1–4** and projected areas from molecular models. **a**, P–A isotherms. **b**, Molecular models of adjacent polymer packing. Top views of two repeating units of polymers **1** and **2** are shown. The calculated length of one repeating unit is $14 \text{ \AA}$, which is used to calculate the area per repeating unit from the modelling.

indicating, as mentioned earlier, that a non-planar conformation similar to the zipper structure is present in solution. The blue-shifted fluorescence λ_{max} of the zipper phase (462 nm) is also similar to the solution value (457 nm). The zipper structure is further supported by the synthesis of polymers with small amounts of ethynylene homo-coupling (that is, diacetylene groups) that disrupt the interlocking structure of polymer chains. A regularly alternating -A-C-A-C-A-C- structure is required to observe clear transitions in the P-A isotherm. Small amounts of an -A-C-C-A-C- imperfection¹⁷ gives a less-featured P-A isotherm. After a monolayer of **2** folds into a multilayer film, a similar behaviour to that of **1** is observed, with the formation of an aggregation peak at 470 nm.

Langmuir monolayers of polymer **3** appear to assemble into the zipper structure before compression, and have an extrapolated area per repeating unit of 150 Å² that is much smaller than that of **1** and **2**. This zipper structure is promoted by the phenyl ring of building block D, which, when orientated normal to the surface, minimizes contact between the two hydrophobic sidechains and the water. The shape and λ_{max} of **3**'s uncompressed monolayer absorption and fluorescence spectra are consistently the same as those in solution. As **3** is compressed, one of the moderately hydrophilic triethylene-oxide groups from building block B is removed from the aqueous phase, and the film transforms from a zipper to an edge-on structure between 150 and 100 Å². This structural transition is probably facilitated by favourable interpolymer dipolar interactions

between these polar side chains in the edge-on structure.

In the edge-on structure polymer **3** is co-facially arranged, and—as we have previously established— π -aggregation between main chains generates a new aggregation band in the UV-vis (Fig. 3a) spectra. This new narrow red-shifted band with a distinct shoulder is characteristic of a π -aggregated state with lower quantum yields, and should be distinguished from the red-shift owing to planarization alone¹⁸ observed for **1** and **2**. The film in an edge-on structure is aggregated before film collapse; therefore no new band arises upon folding into a multilayer. As the aggregation band in the UV-vis spectra grows the fluorescence is quenched, and when completely aggregated the quantum yield drops to 16% of the initial non-aggregated value (Fig. 3b). Before the folding of the monolayer into multilayers, the transition from the zipper to the edge-on structure is completely reversible. Therefore, releasing the surface pressure de-aggregates the film, and the initial absorbance and fluorescence spectra are re-established.

Polymer **4** has two edge-on building blocks, and forms the edge-on structure even before compression. Molecular modelling for **3** and **4** in an edge-on structure predicts an area per repeating unit of 64 Å², which is consistent with the value of 60 Å² from the P-A isotherm. As the film is compressed, the fluorescence intensity gradually decreases because the reduced distance between co-facially arranged polymer main chains promotes self-quenching^{19,20}. The area per repeating unit at which the monolayer folds into

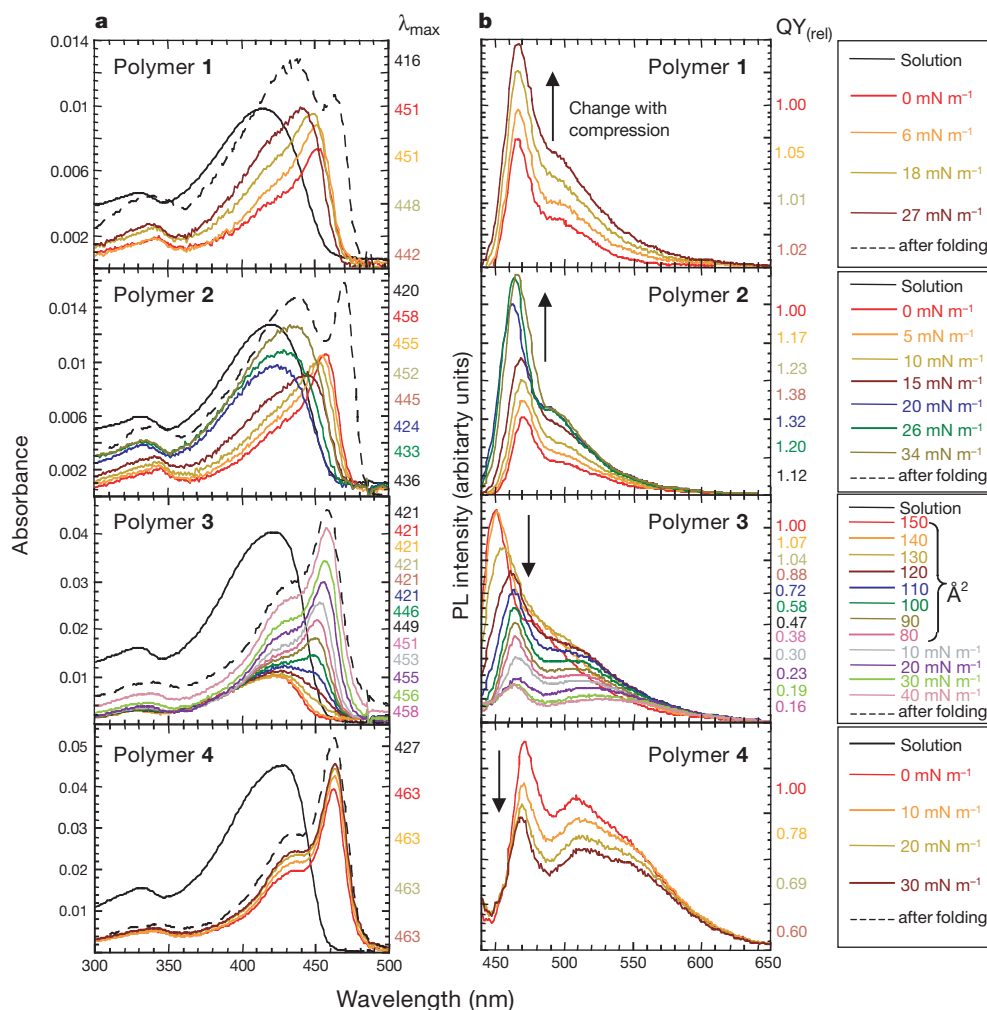


Figure 3 *In situ* UV-vis. and fluorescence spectra of Langmuir monolayers of polymers **1–4** during cycles of compressions and expansions. Polymer solutions were prepared in chloroform. **a**, *In situ* UV-vis. spectra and λ_{max} at various surface pressures. **b**, *In situ*

fluorescence spectra and relative quantum yields (QY_{rel}) at various surface pressures and area per repeating units (relative quantum yield at 0 mN m⁻¹ = 1). Excitation wavelength was 420 nm.

multilayers is $50 \text{ \AA}^2$, the same value as for polymer **3**. This observation also supports the transition of **3** from the zipper to the edge-on. The aggregate emission bands are also identical with those of highly compressed **3**. The aggregate features are present at an area per repeating unit of $60 \text{ \AA}^2$. This value and a repeat unit length of $14 \text{ \AA}$ provide an interpolymer distance of $4.3 \text{ \AA}$, provided that the phenyl rings are standing perpendicular to the air–water interface. At higher pressures the emission is further quenched, and at the maximum compression of monolayers of polymers **3** and **4** the area per repeating unit is $50 \text{ \AA}^2$, which provides an interpolymer spacing of $3.6 \text{ \AA}$. The red-shifted absorption band implies that the aggregation is probably J-type. But because we have no knowledge of the registry of polymer chains with respect to each other, we cannot confirm whether the aggregation is of type H (in phase) or type J (out of phase).

To verify the nature of the emission peak at 471 nm and the sharp absorption peak at 463 nm for polymer **4**, we conducted non-solvent induced aggregation studies. Polymer **4** was dissolved in chloroform, and then UV–vis. and fluorescence spectra were taken as methanol, a non-solvent, was gradually added to the solution. Above equal amounts of chloroform and methanol, the polymer aggregates and absorption spectra display a pseudo-isosbestic point. (An isosbestic point implies that there are two different interconverting species.) Because of the limited solubility of **4**, it was technically difficult to conduct the experiment at a constant molar concentration, which is a necessary condition for a true isosbestic point. The absorption λ_{max} of the aggregate is 460 nm , close to the 463 nm of **4** at the air–water interface. The non-aggregated emission peak at 452 nm is continuously suppressed with added methanol until the spectrum has a main peak at 475 nm and two other peaks at about 510 and 545 nm . The final fluorescence spectrum is almost identical with the fluorescence spectrum of **4** in the edge-on structure at the air–water interface.

Excitation spectra of a solution in which an aggregated and a non-aggregated phase coexist have been collected by monitoring the emissions at 450 , 475 , 510 and 545 nm . The emission peak at 450 nm corresponds to the non-aggregated phase, as its excitation spectrum displays a maximum at 420 nm . The emission peaks at 475 , 510 and 545 nm all have the same excitation spectra with peaks at 420 and 460 nm . Therefore, these three peaks are the result of polymer aggregates. The presence of the 420-nm peak in the excitation spectra is due to energy transfer from the higher-energy non-aggregated phase to the aggregated phase. We can discount the possibility of an excimer because we clearly see a ground-state absorption that is responsible for the new emission bands²¹. An excimer is an excited state complex, and should not result in new absorption bands. The lifetime of the non-aggregated band at

450 nm is 0.5 ns , and those of the other bands at 475 , 510 and 545 nm are $0.1\text{--}0.2 \text{ ns}$. The lifetimes of excimers are often much longer than intrachain excitons²¹, which also supports the fact that the longer-wavelength bands are not excimer peaks.

Selected UV–vis. spectra of **1–4** from Fig. 3a are compared in Fig. 4. These spectra show the self-consistency of our results that have allowed us to unambiguously assign the structures to different spectral attributes. Polymers **1** and **2** have the same face-on structure at 0 mN m^{-1} ; then mechanical pressure at 20 mN m^{-1} induces transition of **2** into the zipper structure—which **3** has initially at 0 mN m^{-1} . Mechanical compression of **3** causes the transition into the edge-on structure with a co-facial π -aggregation band that **4** has at 0 mN m^{-1} .

The combination of unique surfactant design and mechanical interrogation at the air–water interface have allowed us to show the interrelationships between the conformation of a single chain and the interpolymer interactions and a conjugated polymer's absorption and emission properties. Although these materials have been studied extensively^{1–4,8,10–11,14}, the intrinsic properties of isolated chains and ordered aggregates have yet to be established. The influence of structure on the charge- and energy-transporting properties of conjugated polymers could also be addressed by extensions of the methods presented here. □

Methods

A Nima 601 M model Langmuir–Blodgett trough with a window was used for the studies. The *in situ* UV–vis. spectra were obtained on a Hewlett–Packard 8453 diode array spectrophotometer or a Cary 50 Scan with fibre optics. The *in situ* fluorescence studies were conducted with a SPEX Fluorolog-72 fluorometer equipped with fibre optics. All the spectroscopic measurements on Langmuir films were conducted through optical fibres. Molecular modelling was performed using the Spartan 5.0 molecular modelling program (Wavefunction Inc, Irvine, California) running on a SGI O2 (R12000) workstation. These calculations are intended to give good estimates of the areas per repeating unit and required some geometric constraints that simulate the organizational preferences at the air–water interface.

Received 8 February; accepted 13 April 2001.

1. Nguyen, T.-Q., Wu, J., Doan V., Schwartz, B. J. & Tolbert, S. H. Control of energy transfer in oriented conjugated polymer-mesoporous silica composites. *Science* **288**, 652–656 (2000).
2. Shi, Y., Liu, J. & Yang, Y. Device performance and polymer morphology in polymer light emitting diodes: The control of thin film morphology and device quantum efficiency. *J. Appl. Phys.* **87**, 4254–4263 (2000).
3. Nguyen, T.-Q., Martini, I. B., Liu, J. & Schwartz, B. J. Controlling interchain interactions in conjugated polymers: The effects of chain morphology on exciton-exciton annihilation and aggregation in MEH-PPV films. *J. Phys. Chem. B* **104**, 237–255 (2000).
4. Sirringhaus, H. *et al.* Two-dimensional charge transport in self-organized, high-mobility conjugated polymers. *Nature* **401**, 685–688 (1999).
5. Lucht, B. L., Mao, S. S. H. & Tilley, T. D. A zirconocene-coupling route to substituted poly(p-phenylenedienylene)s: Band gap tuning via conformational control. *J. Am. Chem. Soc.* **120**, 4354–4365 (1998).
6. Cornil, J., dos Santos, D. A., Crispin, X., Silbey, R. & Brédas, J. L. Influence of interchain interactions on the absorption and luminescence of conjugated oligomers and polymers: A quantum-chemical characterization. *J. Am. Chem. Soc.* **120**, 1289–1299 (1998).
7. Yang, J.-S. & Swager, T. M. Fluorescent porous polymer films as TNT chemosensors: Electronic and structural effects. *J. Am. Chem. Soc.* **120**, 11864–11873 (1998).
8. Buvat, P. & Hourquebie, P. Enhanced infrared properties of regioregular poly(3-alkylthiophenes). *Macromolecules* **30**, 2685–2692 (1997).
9. Jenekhe, S. A. & Osaheni, J. A. Excimers and exciplexes of conjugated polymers. *Science* **265**, 765–768 (1994).
10. Gettinger, C. L., Heeger, A. J., Drake, J. M. & Pine, D. J. A photoluminescence study of poly(phenylene vinylene) derivatives—The effect of intrinsic persistence length. *J. Chem. Phys.* **101**, 1673–1678 (1994).
11. Heun, S. *et al.* Conformational effects in poly(p-phenylene vinylene)s revealed by low-temperature site-selective fluorescence. *J. Phys. Condens. Matter* **5**, 247–260 (1993).
12. Kim, J., McHugh, S. K. & Swager, T. M. Nanoscale fibrils and grids: Aggregated structures from rigid-rod conjugated polymers. *Macromolecules* **32**, 1500–1507 (1999).
13. Josefowicz, J. Y. *et al.* Structure of Langmuir–Blodgett films of disk-shaped molecules determined by atomic force microscopy. *Science* **260**, 323–326 (1993).
14. Zhang, Q. T. & Tour, J. M. Imine-bridged planar poly(phenylenethiophene)s and polythiophenes. *J. Am. Chem. Soc.* **119**, 9624–9631 (1997).
15. Levitsky, I. A., Kim, J. & Swager, T. M. Energy migration in a poly(phenylene ethynylene): determination of interpolymer transport in anisotropic Langmuir–Blodgett films. *J. Am. Chem. Soc.* **121**, 1466–1472 (1999).
16. Turro, N. J. *Modern Molecular Photochemistry* 183 (University Science Books, Sausalito, California, 1991).
17. Kloppenburg, L., Jones, D. & Bunz, U. H. F. High molecular weight poly(p-phenyleneethynylene)s by alkyne metathesis utilizing “instant” catalysts: A synthetic study. *Macromolecules* **32**, 4194–4203 (1999).

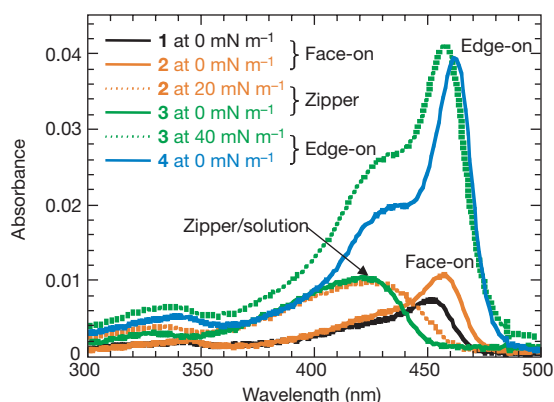


Figure 4 Selected UV–vis. spectra of polymers **1–4** in different structures (face-on, zipper, and edge-on). These data show the interrelationship between the conformation and interpolymer interactions.

18. Miteva, T., Palmer, L., Kloppenburg, L., Neher, D. & Bunz, U. H. F. Interplay of thermochromicity and liquid crystalline behavior in poly(*p*-phenyleneethynylene)s: π - π interactions or planarization of the conjugated backbone? *Macromolecules* **33**, 652–654 (2000).
19. McQuade, D. T., Kim, J. & Swager, T. M. Two-dimensional conjugated polymer assemblies: Interchain spacing for control of photophysics. *J. Am. Chem. Soc.* **122**, 5885–5886 (2000).
20. Kim, J., McQuade, D. T., McHugh, S. K. & Swager, T. M. Ion-specific aggregation in conjugated polymers: Highly sensitive and selective fluorescent ion chemosensors. *Angew. Chem. Int. Edn Engl.* **39**, 3868–3872 (2000).
21. Cronwell, E. Excimer formation and luminescence in conducting polymers. *Trends Polym. Sci.* **5**, 218–222 (1997).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank Z. Zhu for providing 1,2-didodecyloxy-3,6-diethynylbenzene (building block D), and A. Rose for the lifetime measurements. This work was supported by the Office of Naval Research and Draper Laboratory.

Correspondence and requests for materials should be addressed to T.M.S. (e-mail: tswager@mit.edu).

Biodegradation of oil in uplifted basins prevented by deep-burial sterilization

A. Wilhelms*, S. R. Larter†, I. Head†, P. Farrimond†, R. di-Primio* & C. Zwach*

* Norsk Hydro Research Centre, PO 7190, 5020 Bergen, Norway

† Fossil Fuels and Environmental Geochemistry Postgraduate Institute (NRG), University of Newcastle, Newcastle upon Tyne NE1 7RU, UK

Biodegradation of crude oil by bacterial activity—which has occurred in the majority of the Earth's oil reserves¹—is known to reduce greatly the quality of petroleum in reservoirs². For economically successful prospecting for oil, it is therefore important to understand the processes and conditions in geological formations that lead to oil biodegradation. Although recent studies speculate that bacterial activity can potentially occur up to temperatures as high as 150 °C (refs 3, 4), it is generally accepted that effective petroleum biodegradation over geological time-scales generally occurs in reservoirs with temperatures below 80 °C (ref. 2). This appears, however, to be at odds with the observation that non-degraded oils can still be found in reservoirs below this temperature. Here we compile data regarding the extent of oil biodegradation in several oil reservoirs, and find that the extensive occurrence of non-biodegraded oil in shallow, cool basins is restricted to those that have been uplifted from deeper, hotter regions of the Earth. We suggest that these petroleum reservoirs were sterilized by heating to a temperature around 80–90 °C during deep burial, inactivating hydrocarbon-degrading organisms that occur in the deep biosphere. Even when such reservoirs are subsequently uplifted to much cooler regions and filled with oil, degradation does not occur, implying that the sterilized sediments are not recolonized by hydrocarbon-degrading bacteria.

Biodegradation of oil leads to a systematic decrease in paraffin content and an increase in oil density, sulphur content, acidity and viscosity², with negative economic consequences for oil production and refining operations. So prediction of the degree of biodegradation of oil is important for assessing the risk of finding degraded oils in an exploration target before drilling. The details and rates of the processes involved in crude-oil degradation are still poorly understood, but the central role of bacteria in subsurface petroleum degradation is accepted. Although there is abundant evidence for

the presence of active bacteria deep (>1 km) in the Earth's crust^{3,5–7}, and a view among biologists that life in deep sediments may occur even up to 150 °C (refs 3, 4), there is a general view today among petroleum geoscientists that biodegradation in reservoirs ceases around 75–80 °C (refs 2, 8) in the zone of thermophilic organisms.

Hyperthermophilic organisms grow best above 80 °C, and have been reported to live at temperatures of up to 113 °C (ref. 9), but often display no growth below 60 °C (ref. 4). Most reports of hyperthermophilic activity are limited to environments rich in reduced electron donors, electron acceptors and inorganic nutrients necessary for biosynthesis—environments typically associated with shallow sediment or near-surface hydrothermal activity¹⁰, where high levels of metabolic activity can be supported. Although petroleum reservoirs are rich in reduced organic electron donors (hydrocarbons), most petroleum reservoirs and deep aquifers are nutrient depleted¹¹ and recent studies indicate that biodegradation of petroleum in deep reservoirs is occurring at very slow net rates, typically consuming around 10^{–6} mmol oil per litre per day (ref. 12). These are similar to the respiration rates suggested for other deep sediments, and several orders of magnitude slower than respiration rates in anaerobic laboratory or near-surface sediment environments^{11,13,14}. As electron donor (oil) supply is not limiting, degradation is presumably nutrient limited. Under these stressed, nutrient-limited conditions, it is unlikely that the temperature extremes survived by organisms in nutrient-rich hydrothermal environments are relevant. Furthermore, the empirical upper temperature limit of 80 °C observed for petroleum biodegradation in deep reservoirs is considerably lower than temperatures quoted for the denaturation and decomposition of many bacterial and archaeal proteins and nucleic acids^{9,15}; this implies that active hyperthermophilic hydrocarbon-degraders are probably not present in petroleum reservoirs. Indeed, the single report of a hyperthermophilic archaeon that grew in enrichment cultures with crude oil as sole source of carbon¹⁶ did not provide unequivocal evidence that the hyperthermophiles isolated could degrade hydrocarbons.

Although the occurrence of hyperthermophiles in petroleum reservoirs is now well documented^{17–21}, only four isolates capable of growth at greater than 90 °C have been reported^{16–19}. In the most extreme of these cases (growth at 102 °C), evidence was presented that suggested that the organisms concerned may have originated from hydrothermal vent systems and were introduced with injected sea water¹⁶. The other cases where hyperthermophiles have been isolated are also sea-water-flooded reservoirs¹⁹. For the most part, microbial enrichments from petroleum reservoir samples conducted at temperatures above 85 °C have been unsuccessful^{19,21}, with increasing success of enrichment cultures with decreasing growth temperature (reaching 77% ($n = 26$) at temperatures below 60 °C)¹⁹. We also note that the only thermophilic anaerobic bacterium for which hydrocarbon degradation has been unequivocally demonstrated (an anaerobic sulphate-reducer) has an optimum growth temperature of only 60 °C (ref. 22). It has also been found that enrichment cultures supplied with crude oil as the sole source of organic carbon at temperatures greater than 85 °C failed to produce cultures of putative hydrocarbon-degrading Archaea although hyperthermophiles that grow at temperatures up to 102 °C could be isolated¹⁶. Taken together, data from most reports of Bacteria and Archaea isolated from petroleum reservoirs support the notion that the organisms which inhabit these reservoirs do not thrive in the upper temperature ranges for hyperthermophiles (up to 113 °C), and show little indication of hydrocarbon degrading capacity above 80 °C. We realize, however, that any interpretation must be viewed cautiously (as most reports provide little information on failed enrichments, and recent experience in defining the limits of life warn of inherent difficulty in definitively excluding life processes from even severe environments).

Biodegraded oils are common in cool (<80 °C), shallow reservoirs in currently subsiding basins such as those in the North Sea,

the Gulf of Mexico and the Niger delta, or where very near surface exposure of oil occurs—as with tar sands¹ or oil seeps. However, some cool, shallow reservoirs contain non-degraded petroleum, and these are of particular interest for petroleum exploration because of their greater commercial value and ease of exploitation. In our experience, such non-degraded shallow oils are chiefly found either in reservoirs with a very recent oil charge or, more interestingly, in reservoirs that have been uplifted from their maximum burial depth. Examples of basins containing light non-degraded oils in currently shallow cool reservoirs that have been exposed to substantially higher temperatures in the past occur in the Barents Sea (Norway), Wessex and East Midlands basins (UK), Michigan, San Juan, Anadarko and Illinois basins (USA), Murzuk basin (Libya), Zagros folded zone (Iraq), Cooper-Eromanga basins (Australia) and the Papua fold and thrust belt (see Supplementary Information). Although late oil recharge into these uplifted reservoirs may explain the observed low levels of net biodegradation, we consider this process unlikely to have been significant in most cases. This is because the hydrocarbon-generating source rocks cool when the basins are uplifted, and primary oil charging ceases.

Comparison of levels of biodegradation in similar petroleum systems that have very different reservoir temperature histories before oil charging indicates that uplifted and continuously subsiding basins show quite different biodegradation characteristics (Fig. 1). In the uplifted Barents Sea (Norway) and Wessex (UK) basins, oils are currently found in reservoirs over depth ranges of 1,000–2,000 m and 500–1,600 m, respectively, with some of these oils being found at temperatures below 40 °C. These oils are uniformly non-degraded (Fig. 2). In contrast, oils in the currently subsiding Viking graben found in reservoirs between 1,000 and 2,000 m depth show variable pristane/*n*-heptadecane ratios ranging from <1.0 (non-degraded) to over 10 (Fig. 2). This indicates a wide range of degradation, related to the relative timing of fresh oil charge, and variable remixing of fresh and degraded oils in the reservoirs^{8,23}. All three petroleum systems are similar in that the source rocks are Jurassic marine shales of similar mineralogy, with

similar type II kerogen source organic matter (Draupne Formation²⁴, Hekkingen Formation²⁵ and the Lias²⁶) producing broadly similar initial gross oil compositions in the reservoirs.

The Viking graben reservoirs have had a relatively continuous subsidence history, marked by only minor uplift events. All the reservoirs are at present at maximum temperature, and the reservoir temperature in these shallow degraded Tertiary reservoirs has never exceeded 80 °C (ref. 27). Basin modelling indicates that oil charging of the traps occurred from the Late Cretaceous, but most traps have had oil charges as recent as 5 Myr ago (ref. 28). Generally well defined trends of depth versus degradation observed regionally in the Viking graben⁸ suggest that degradation has occurred close to current reservoir depths. Figure 2 also shows data for oils from Jurassic reservoirs in the Tampen Spur area of the North Sea which contain essentially the same type of oil as found in the Tertiary reservoirs of the Viking graben but are at their maximum depths of greater than 2,000 m and at temperatures of over 80 °C. These oils are non-degraded.

The reservoirs in the Barents Sea and Wessex basins have had a more turbulent burial history, including extensive uplift. In the case of the Barents Sea, major oil charging occurred in the Cretaceous followed by uplift of the reservoirs by >1.5–2 km in the early Tertiary^{25,28}. Maximum temperatures experienced by the currently shallowest reservoirs were around 85 °C, with the deeper reservoirs encountering higher temperatures. The Wessex basin has a similar history. Oil charging from the Lias occurred in the Cretaceous²⁶, and was followed in late Cretaceous time by uplift of approximately 2 km (refs 29, 30). None of the reservoir oils at the Wytch Farm or Kimmeridge Bay oilfields in this basin show any signs of degradation, despite the oils being in cool reservoirs as shallow as 500 m. Maximum temperatures of the reservoirs at Wytch Farm and Kimmeridge Bay oilfields are indicated to have been around 100 °C and 90 °C (ref. 29) respectively. It is probable that the current oil charge at Wytch Farm has been there since the Cretaceous²⁶, and that at Kimmeridge Bay oilfield since soon after the Tertiary uplift³⁰.

Accordingly, reservoir sequences that have been exposed to

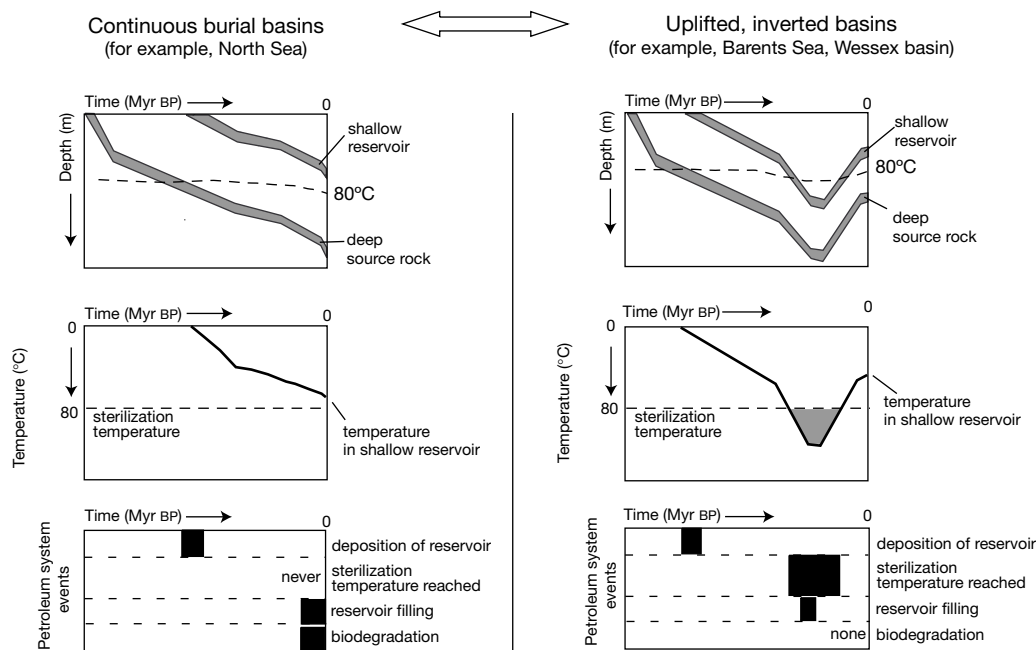


Figure 1 Comparison of continuously subsiding and uplifted sedimentary basins. The former are represented by the Viking graben, Tampen Spur and North Sea basins, and the latter by the Barents Sea and Wessex basins. Charts show schematic burial history (top), reservoir temperature history (middle), and petroleum system events (bottom), illustrating

key differences with respect to biodegradation in petroleum reservoirs. Biodegradation of petroleum occurs only in sedimentary units that have not been exposed to temperatures exceeding 80 °C ('palaeosterilization') before oil charging.

temperatures in excess of $\sim 90^\circ\text{C}$ before the current oil charge seem to be no longer capable of supporting petroleum biodegradation, even though the reservoir is far from chemical equilibrium and thus chemical energy to support life is present (that is, both hydrocarbons and potential oxidants including excess water¹³ coexist). It seems most unlikely that all of the Triassic to Upper Jurassic fluvial to marine reservoir sediment systems of the Wessex and Barents Sea basins were devoid of hydrocarbon-degrading bacteria initially, because comparable reservoir settings in the Viking graben do have degraded oils. We suggest that the uplifted reservoirs do not contain degraded oils because the hydrocarbon-degrading bacteria of the reservoirs were inactivated by deep burial before the oil charge entered—'palaeosterilization'.

It now seems likely that most sediments cooler than the maximum temperature boundary of the deep biosphere contain microorganisms^{3,7,31}. If 'palaeosterilization' occurs, it appears that once destroyed, effective bacterial and archaeal communities (or at the very least hydrocarbon-degrading bacterial and archaeal communities) are not re-established in deep reservoirs. This is true even in the Wessex and Barents Sea basins that had large surface exposure of terrains ostensibly suitable for recharge of meteoric waters, nutrients and microorganisms. These observations indicate that both surface recharge of fluids or organisms, and migration of organisms in the deep subsurface environment of petroleum basins, may be insignificant. Thus, in deep reservoirs where degradation does occur, microorganisms must have already been in place during slow burial, and have survived and probably evolved on geological timescales. This implies that the microbial flora of deep reservoirs may be isolated biospheres, descendants of ancient lineages responding in an isolated manner to evolution over millions of years, as suggested for other sediments³¹.

Our reservoir palaeo-temperature estimates suggest that the base of the hydrocarbon-degrading biosphere is at temperatures lower than 90°C , in agreement with much oilfield data and the empirical

rule used by petroleum geologists that biodegraded oils are only found in reservoirs cooler than 80°C . Our proposed 'palaeosterilization' model does not provide information before drilling about the degree of oil biodegradation to be expected in continuously subsiding basins. However, it does explain extensive, shallow occurrences of non-degraded oil in many uplifted reservoirs that are at equal or lower temperatures than those at which degraded oils are found in other basins (see Supplementary Information). This simple model, if further validated, implies that some very shallow reservoirs, which have been avoided because of the risk of petroleum biodegradation, would now represent much lower risk, low cost, exploration targets if they have been buried to temperatures in excess of $\sim 80\text{--}90^\circ\text{C}$. □

Received 7 September 2000; accepted 2 May 2001.

1. Roadifer, R. E. in *Exploration for Heavy Crude Oil and Natural Bitumen* (ed. Meyer, R. F.) 3–23 (AAPG Studies in Geology 25, American Association of Petroleum Geologists, Tulsa, 1987).
2. Connan, J. in *Advances in Petroleum Geochemistry* Vol. 1 (eds Brooks, J. & Welte, D. H.) 299–335 (Academic Press, London, 1984).
3. Parkes, R. J. et al. Deep bacterial biosphere in Pacific Ocean sediments. *Nature* **371**, 410–413 (1994).
4. Stetter, K. O., Fiala, G., Huber, G., Huber, R. & Seeger, A. Hyperthermophilic microorganisms. *FEMS Microbiol. Rev.* **75**, 117–124 (1990).
5. ZoBell, C. E. The role of bacteria in the formation and transformation of petroleum hydrocarbons. *Science* **102**, 364–369 (1945).
6. Jobson, A., Cook, F. D. & Westlake, D. W. S. Microbial utilization of crude oil. *Appl. Microbiol.* **23**, 1082–1089 (1972).
7. Pedersen, K. Exploration of deep intraterrestrial microbial life; current perspectives. *FEMS Microbiol. Lett.* **185**, 9–16 (2000).
8. Barnard, P. C. & Bastow, M. A. in *Petroleum Migration* (eds England, W. A. & Fleet, A. J.) 167–190 (Special Publication No. 59, The Geological Society, London, 1991).
9. Blöchl, E. et al. *Pyrolobus fumarii*, gen. and sp. nov., represents a novel group of archaea, extending the upper temperature limit for life to 113 degrees C. *Extremophiles* **1**, 14–21 (1997).
10. Stetter, K. O. Hyperthermophilic prokaryotes. *FEMS Microbiol. Rev.* **18**, 149–158 (1996).
11. Murphy, E. M. & Schramke, J. A. Estimation of microbial respiration rates in groundwater by geochemical modelling constrained with stable isotopes. *Geochim. Cosmochim. Acta* **62**, 3395–3406 (1998).
12. Larter, S. R. et al. in *Conference Abstracts, Proc. GeoCanada 2000* (eds Bachu, S. and Esslinger, P.) Abstract 120, 1–4 (CSPG/GAC, Calgary, Alberta, 2000).
13. Zengler, K., Richnow, H. H., Rossella-Mora, R., Michaelis, W. & Widdel, F. Methane formation from long-chain alkanes by anaerobic microorganisms. *Nature* **401**, 266–269 (1999).
14. Chapelle, F. H. & Lovley, D. R. Rates of microbial-metabolism in deep coastal-plain aquifers. *Appl. Environ. Microbiol.* **56**, 1865–1874 (1990).
15. Daniel, R. M. & Cowan, D. A. Biomolecular stability and life at high temperatures. *Cell Mol. Life Sci.* **57**, 250–264 (2000).
16. Stetter, K. O. et al. Hyperthermophilic Archaea are thriving in deep North Sea and Alaskan oil reservoirs. *Nature* **365**, 743–745 (1993).
17. Takahata, Y., Nishijima, M., Hoaki, T. & Maruyama, T. Distribution and physiological characteristics of hyperthermophiles in the Kubiki oil reservoir in Niigata, Japan. *Appl. Environ. Microbiol.* **66**, 73–79 (2000).
18. L'Haridon, S., Reysenbach, A. L., Glénat, P., Prieur, D. & Jeanthon, C. Hot subterranean biosphere in a continental oil reservoir. *Nature* **377**, 223–224 (1995).
19. Grassia, G. S., McLean, K. M., Glénat, P., Bauld, J. & Sheehy, A. J. A systematic survey for thermophilic fermentative bacteria and archaea in high temperature petroleum reservoirs. *FEMS Microbiol. Ecol.* **21**, 47–58 (1996).
20. Nilsen, R. K., Beeder, J., Thorstenson, T. & Torsvik, T. Distribution of thermophilic sulfate reducers in North Sea oil field reservoir waters and oil reservoirs. *Appl. Environ. Microbiol.* **62**, 1793–1798 (1996).
21. Orphan, V. J., Taylor, L. T., Hafenbradl, D. & Delong, E. F. Culture-dependent and culture-independent characterization of microbial assemblages associated with high-temperature petroleum reservoirs. *Appl. Environ. Microbiol.* **65**, 700–711 (1999).
22. Rueter, P. et al. Anaerobic oxidation of hydrocarbons in crude oil by new types of sulphate-reducing bacteria. *Nature* **372**, 455–458 (1994).
23. Horstad, I. & Larter, S. R. Petroleum migration, alteration, and remigration within Troll Field, Norwegian North Sea. *Bull. Am. Assoc. Petrol. Geol.* **81**, 222–248 (1997).
24. Cornford, C. in *Introduction to the Petroleum Geology of the North Sea* (ed. Glennie, K. W.) 294–361 (Blackwell Scientific Publications, Oxford, 1990).
25. Linjordet, A. & Grung Olsen, R. in *Giant Oil and Gas Fields of the Decade 1978–1988* (ed. Halbouty, M. T.) 349–370 (American Association of Petroleum Geologists, Tulsa, Oklahoma, 1992).
26. Cornford, C., Christie, O., Endresen, U., Jensen, P. & Myhr, M. B. Source rock and seep oil maturity in Dorset, southern England. *Org. Geochem.* **13**, 399–409 (1988).
27. Schroeder, F. W. & Sylta, Ø. in *Basin Modelling: Advances and Applications* (eds Doré, A. G. et al.) 469–484 (Vol. 3, NPF Special Publication, Elsevier, Amsterdam, 1993).
28. Theis, N. J., Nielsen, H. H., Sales, J. K. & Gail, G. J. in *Basin Modelling: Advances and Applications* (eds Doré, A. G. et al.) 433–444 (Vol. 3, NPF Special Publication, Elsevier, Amsterdam, 1993).
29. Bray, R. J., Duddy, I. R. & Green, P. F. in *Development, Evolution and Petroleum Geology of the Wessex Basin* (ed. Underhill, J. R.) 199–213 (Special Publication 133, Geological Society, London, 1998).
30. Evans, J., Jenkins, D. & Gluyas, J. in *Development, Evolution and Petroleum Geology of the Wessex Basin* (ed. Underhill, J. R.) 407–413 (Special Publication 133, Geological Society, London, 1998).
31. Frederickson, J. K. et al. Microbial community structure and biogeochemistry of Miocene subsurface sediments—implications for long-term microbial survival. *Mol. Ecol.* **4**, 619–626 (1995).

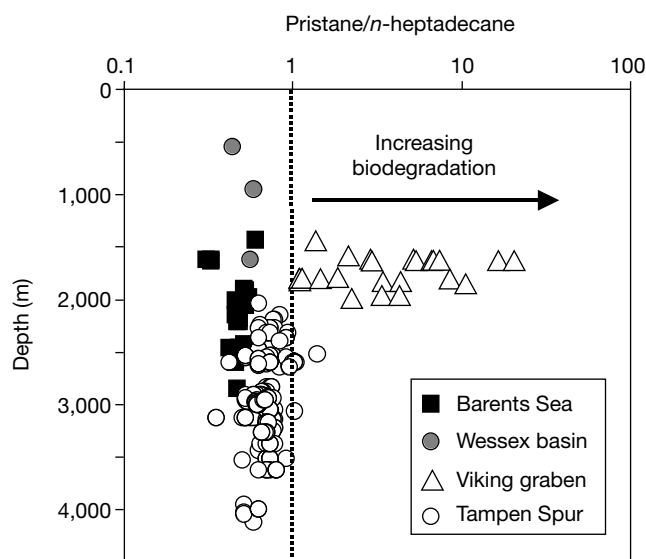


Figure 2 Degree of oil biodegradation, assessed by the pristane/*n*-heptadecane ratio, as a function of current reservoir depth for oils from four different settings. The data derive from Jurassic and Triassic reservoirs of the Wessex basin, Jurassic reservoirs of the Barents Sea, Palaeocene and Eocene reservoirs of the Viking graben and Jurassic reservoirs of the Tampen Spur and Horda platform, and were obtained by gas chromatographic analysis of whole oils or separated saturated hydrocarbon fractions. Values greater than ~ 1 represent biodegraded oils. The Viking graben and Tampen Spur oils (North Sea) are from continuously subsiding reservoirs (empty symbols), whilst the Barents Sea and Wessex basin reservoirs were uplifted and recharged with oil after deeper burial (filled symbols).

Supplementary information is available from Nature's World-Wide Web site or as paper copy from the London editorial office of *Nature*.

Acknowledgements

We thank Norsk Hydro for support of the Phoenix project, and M. Jones, A. Aplin, N. Telnæs and M. Koopmans for discussions.

Correspondence should be addressed to S.R.L. (e-mail: Steve.Larter@ncl.ac.uk).

Mineral disequilibrium in lavas explained by convective self-mixing in open magma chambers

S. Couch*, R. S. J. Sparks* & M. R. Carroll†

* Department of Earth Sciences, Bristol University, Bristol BS8 1RJ, UK
† INFN and Dipartimento di Scienze della Terra, Università di Camerino, 62032 Camerino, Italy

Characteristic features of many porphyritic andesite and dacite lavas are that they are rich in crystals and display a range of disequilibrium features, including reversely zoned crystals, resorption surfaces, wide ranges of mineral compositions and minerals which are not in equilibrium with the surrounding rock matrix. These features are often interpreted as evidence of the mixing of magmas of contrasting composition, temperature and origin^{1,2}. Here, however, we propose that such features can also be caused by convection within a magma body with a single composition, that is heated from below and cooled from above. We describe petrological observations of andesite lava erupted at the Soufrière Hills volcano, Montserrat, which indicate a heating event and the intermingling of crystals that have very different thermal histories. We present experimental data on a representative groundmass composition of this lava, which indicate that it is difficult to explain the calcic compositions of plagioclase overgrowth rims and microphenocrysts unless parts of the magma were at temperatures much higher than the inferred average temperature. The concept of convective self-mixing allows us to explain the occurrence of compositions of minerals that apparently cannot coexist under equilibrium conditions.

Here we summarize key features of the petrology of the Soufrière Hills andesite^{3,4}, and present new experimental data which further constrains magma genesis. The lava contains 45–55% phenocrysts of plagioclase, hornblende, orthopyroxene, oxide and minor quartz, and microphenocrysts (crystals 100 to 300 μm in length) of plagioclase, clinopyroxene, orthopyroxene and oxide. Variable amounts of microlites (<100 μm in length) are set in a high-silica rhyolitic glass. There are abundant features of disequilibrium and mixing of crystals with contrasted thermal histories. Plagioclases in particular have complex and variable zoning patterns, as is typical of calc-alkaline andesites^{5,6}. Some plagioclase phenocrysts have strongly resorbed sieve-textured zones and rims 30–150 μm in thickness of more calcic composition (65–80% anorthite; An_{65-80}) in comparison to cores (An_{48-58}). Some orthopyroxenes have thin overgrowth rims (up to 25 μm) with high Ca content, a feature consistent with crystallization from higher-temperature magma than the core⁴. Hornblendes range from euhedral, to slightly resorbed, to those completely pseudomorphed by coarse pyroxene–plagioclase–oxide intergrowths^{3,4}. Quartz crystals range from those with resorbed outlines to those with clinopyroxene reaction rims. All variants of texture and mineral composition can be observed in a single thin section. Microphenocrysts of plagioclase vary widely in composition

(cores of An_{48-80} , rims of An_{52-75}) with most being more calcic than the phenocryst cores⁴. Groundmass plagioclase microlites are characteristically more sodic (An_{40-75}) than either overgrowth rims or microphenocrysts.

Figure 1a shows a phase diagram of a representative melt composition of the rhyolitic groundmass found in the andesitic Montserrat bulk composition⁷. The diagram, contoured in An content of plagioclase, illustrates the difficulties of reconciling the observed mineralogy and textures with a simple magmatic evolution. Core compositions of the phenocrysts and the overall stability of amphibole are consistent with magma temperatures in the 840 to 870°C range at $p_{\text{H}_2\text{O}} = 120$ to 140 MPa (refs 9 and 7). Water contents of melt inclusions are estimated to be 4–5 wt%, consistent with water partial pressures of 115–130 MPa (ref. 7). The experimental data (Fig. 1) indicate that the overgrowth rims on plagioclase phenocrysts and microphenocrysts are too calcic to be formed at the inferred magma temperatures, despite the fact that the estimated groundmass composition is relatively calcium-rich. Calcic plagioclase compositions also cannot be explained by isothermal decompression, because plagioclase compositions become increasingly sodic at lower $p_{\text{H}_2\text{O}}$ values. Thus, we conclude that calcic plagioclase overgrowth rims and microlites (> An_{60}) formed at substantially higher temperatures (900–1,000°C) and at still elevated $p_{\text{H}_2\text{O}}$ values (>100 MPa). These inferences are consistent with evidence of high-temperature crystallization, up to 970°C, such as

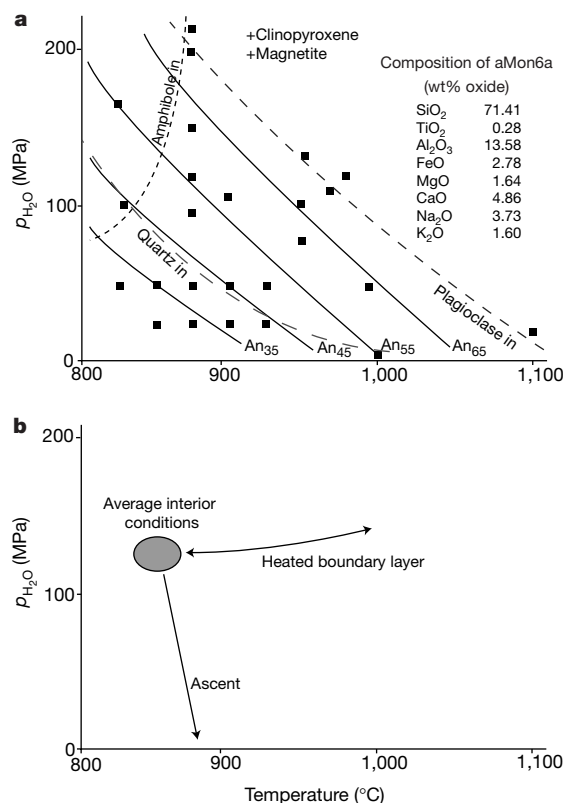


Figure 1 Pressure–temperature diagrams to show conditions of Soufrière Hills magma. **a**, Experimentally derived phase diagram of Montserrat groundmass (aMon6a). The composition was determined by averaging 200 rastered electron-probe analyses of the groundmass. Experiments were run in rapid-quench cold-seal pressure vessels and were water-saturated. Black squares denote experimental conditions; dashed lines denote the boundaries of given phases (plagioclase, quartz, amphibole); solid lines denote contours of plagioclase feldspar anorthite content. All experiments were saturated in clinopyroxene and magnetite. **b**, Schematic diagram of proposed pressure–temperature history of Montserrat magma before eruption. $p_{\text{H}_2\text{O}}$, partial pressure of water.

clinopyroxene microphenocrysts, high-temperature orthopyroxene rims and FeTi oxide data³, but are not easily reconciled with the absence of reaction textures in most hornblendes.

Magma mixing is a possible explanation for the heating and high-temperature crystallization observed in the Montserrat magma. The textures of the mafic inclusions in the andesite show characteristic features of magmatic quenching of hydrous basalt against cooler andesite⁴. However, these inclusions are dominated by fine-grained acicular amphibole with an entirely different composition to the large hornblende phenocrysts in the andesite. Mixing of this magma into the andesite should produce abundant acicular groundmass amphibole xenocrysts, but these are not observed in the andesite.

We propose an alternative explanation based on the concept of a convecting magma body of a single composition heated from below (Fig. 2). Following previous work⁴ we postulate a magma body surrounded by igneous wallrocks of similar mineralogy and composition with emplacement of hydrous basalt at the base before and during the eruption. The thickness of the boundary layer, δ , due to basal heating can be estimated, as previously⁸, from a critical Rayleigh number Ra_c :

$$Ra_c = \frac{g\alpha\Delta T\delta^3\rho}{\kappa\mu} \quad (1)$$

Here g is gravity, α the expansion coefficient, ΔT the temperature difference between the andesite interior and the basalt/andesite interface, ρ the andesite magma density, κ the thermal diffusivity, and μ the dynamic viscosity. For typical values ($\alpha = 2 \times 10^{-4} \text{ K}^{-1}$; $\Delta T \approx 100^\circ\text{C}$; $\rho = 2,600 \text{ kg m}^{-3}$; $\mu = 10^6$ to 10^7 Pa s and $\kappa \approx 8 \times 10^{-7} \text{ m}^2 \text{ s}^{-1}$) of the Soufrière Hills andesite and with $Ra_c = 657.5$ for the free slip boundary between two

fluid layers⁹, boundary layer thicknesses are estimated to be in the range of 1 to 2 m. The estimated value of α takes into account crystal resorption in the heated layer; therefore its value is an order of magnitude greater than the melt thermal expansion coefficient. The boundary layer is a zone of steep temperature, crystal content and viscosity gradients, characterized by crystal resorption as the heated layer grows in thickness. Locally the boundary layer becomes unstable and thus parcels of variably heated andesite are convected up to mix with the cooler main magma body (Fig. 2). The hot boundary layer magma will be quenched against the interior magma to generate overgrowth rims and a higher-temperature groundmass assemblage. If the most intensely heated andesite reached 900–1,000 °C then the precipitation of calcic plagioclase and clinopyroxene microphenocrysts and overgrowth rims from a dacitic melt can be explained (Fig. 1b).

The timescale for mixing can be estimated by scaling arguments noting that $\delta \approx (\kappa t)^{1/2}$, and combining this with equation (1). Timescales of tens of days are estimated for the boundary layer to become unstable. We also estimate, using experimental and theoretical results on Rayleigh–Taylor convection¹⁰, that plumes will be of the order of twice the diameter of the boundary layer thickness (2 to 4 m). The efficiency of the convective mixing can be estimated using values from experimentally derived work^{11,12}. A temperature difference of 100 °C implies a viscosity contrast, U , of approximately 10. For a magma chamber depth of about 1 km the plume Re is around 0.1; thus, convection is laminar with an intermediate mixing efficiency of about 0.5. This complements petrological observations of crystals with varying thermal histories adjacent to each other, and therefore incomplete mixing. Using formulae for plume rise¹⁰ and typical properties of Soufrière Hills andesite magma, we estimate ascent speeds in the range of 5 to 20 m d⁻¹. Thus quenching will be rapid, accounting for the abundance of microphenocrysts and larger microlites. Experimental studies^{13,14} of plagioclase crystallization have shown that there is rapid response to disequilibrium, enabling microlites and overgrowth rims to crystallize over timescales comparable to those associated with convective mixing. For the Soufrière Hills andesite, convective heating from below may be counteracted by assimilation of cooler igneous material from the chamber side, accounting for minor quartz, occasional anomalously low-temperature orthopyroxenes⁴ and entrainment of amphiboles with evidence of hydrothermal alteration from anomalous hydrogen isotope compositions¹⁵. Whether the magma body as a whole heats up or cools down in the long term depends on the relative rates of addition of warmer or cooler magma from the boundary layers of the chamber.

The mechanism of self-mixing in a convecting magma body of single composition heated from below can explain the wide diversity of mineral textures, mineral compositions, diverse apparent temperatures and disequilibria observed in other orogenic igneous rocks. Petrological observations have identified differences in crystal abundance and glass composition of pumices from the 1991 Pinatubo eruption¹⁶. There is no evidence for changes in bulk composition between samples, and crystals in the crystal-poor dacite show evidence of breakage, resorption and crystal dissolution. These features were interpreted as forming in response to pre-heating of the dacite magma by underlying, freshly intruded basaltic magma¹⁶. The dacite lavas of Mount Unzen also show evidence for thermal disequilibrium¹⁷ which can be explained by self-mixing. Hornblende phenocrysts are often surrounded by pargasite rims and plagioclase phenocrysts have rims which are An-rich and include pargasite microlites¹⁷. These textures could not have developed at shallow levels related to degassing and decompression during magma ascent because amphibole is stable only at high water partial pressures. Although the rims and microlites have been interpreted as forming during a magma mixing event¹⁸, we suggest that they could have grown as a result of self-mixing.

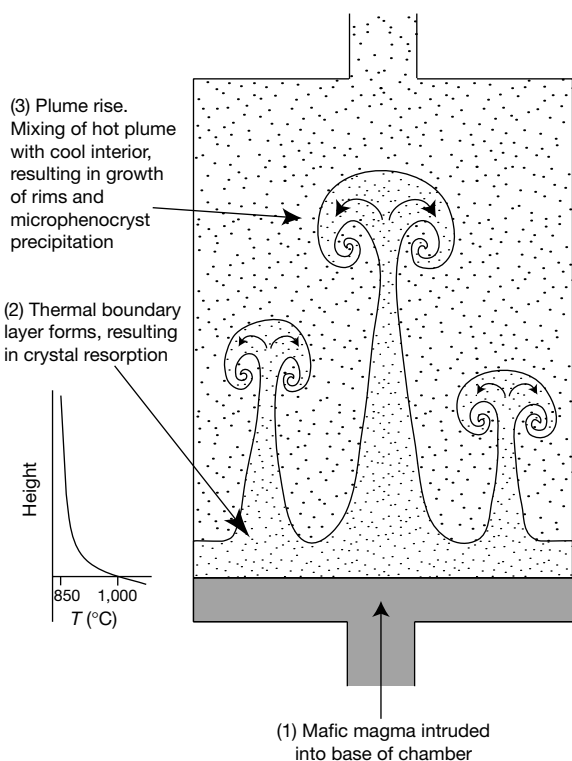


Figure 2 Diagram of self-mixing concept and temperature profile. Hot mafic material is intruded into the base of a more siliceous magma chamber. A layer of hot siliceous magma forms at the boundary. This becomes buoyant and unstable, resulting in the development of convective plumes. The schematic profile shows temperature variation across the thermal boundary layer.

Comparable textural features are observed in granites. For example, feldspar crystals often display rapakivi texture with an alkali feldspar core mantled by a rim of plagioclase feldspar¹⁹. The plagioclase mantle could form from quenching of hotter mafic magma abruptly mixed with felsic magma²⁰; convective self-mixing, however, could also be responsible. Mantled and unmantled feldspar crystals are observed in the same thin section²¹ which could result from localized thermal variations produced by self-mixing. Granites can display, on a local scale, grains of the same phase with contrasting textures, such as large phenocrysts of quartz near to small interstitial grains. There are skeletal and acicular crystal shapes, which are indicative of rapid undercooling and growth, and are consistent with rapid fluctuations in magma conditions²². Orogenic granites are also commonly associated with mafic syn-plutonic intrusions and related mafic inclusions^{23–25}, showing that the necessary conditions for heating and convective self-mixing are widespread.

Convective self-mixing does not require direct involvement of another more mafic magma to generate the complex textures commonly observed in porphyritic orogenic intermediate and silicic volcanic rocks, although mafic magma emplaced at the chamber base is the probable heat source. The concept of self-mixing reconciles otherwise puzzling petrological features, and also supports the notion that magma chambers can convect vigorously. □

Received 21 December 2000; accepted 30 April 2001.

- Heiken, G. & Eichelberger, J. C. Eruptions at Chaos Crags, Lassen Volcanic National Park, California. *J. Volcanol. Geotherm. Res.* **7**, 443–481 (1980).
- Clynne, M. A. A complex magma mixing origin for rocks erupted in 1915, Lassen Peak, California. *J. Petrol.* **40**, 105–132 (1999).
- Devine, J. D. *et al.* Petrologic evidence for pre-eruptive pressure–temperature conditions, and recent reheating, of andesitic magma erupting at the Soufrière Hills Volcano, Montserrat, WI. *Geophys. Res. Lett.* **25**, 3669–3672 (1998).
- Murphy, M. D., Sparks, R. S. J., Barclay, J., Carroll, M. R. & Brewer, T. S. Remobilisation of andesite magma by intrusion of mafic magma at the Soufrière Hills Volcano, Montserrat, West Indies. *J. Petrol.* **41**, 21–42 (2000).
- Kawamoto, T. Dusty and honeycomb plagioclase: indicators of processes in the Uchino stratified magma chamber, Izu Peninsula, Japan. *J. Volcanol. Geotherm. Res.* **48**, 191–208 (1992).
- Tsueh, Y. A. Dissolution kinetics of plagioclase in the melt of the system diopside–albite–anorthite, and origin of dusty plagioclase in andesites. *Contrib. Mineral. Petrol.* **89**, 1–16 (1985).
- Barclay, J. *et al.* Experimental phase equilibria constraints on pre-eruptive storage conditions of the Soufrière Hills magma. *Geophys. Res. Lett.* **25**, 3437–3440 (1998).
- Snyder, D. Thermal effects of the intrusion of basaltic magma into a more silicic magma chamber and implications for eruption triggering. *Earth Planet. Sci. Lett.* **175**, 257–273 (2000).
- Turcotte, D. L. & Schubert, G. *Geodynamics: Applications of Continuum Physics to Geological Problems* (Wiley, New York, 1982).
- Whitehead, J. A. & Luther, D. S. Dynamics of laboratory diapir and plume models. *J. Geophys. Res.* **80**, 705–717 (1975).
- Jellinek, A. M., Kerr, R. C. & Griffiths, R. W. Mixing and compositional stratification produced by natural convection, 1. Experiments and their application to Earth's core and mantle. *J. Geophys. Res.* **104**, 7183–7201 (1999).
- Jellinek, A. M. & Kerr, R. C. Mixing and compositional stratification produced by natural convection, 2. Applications to the differentiation of basaltic and silicic magma chambers and komatiite lava flows. *J. Geophys. Res.* **104**, 7203–7218 (1999).
- Swanson, S. E. Relation of nucleation and crystal-growth rate to the development of granitic textures. *Am. Mineral.* **62**, 966–978 (1977).
- Muncill, G. E. & Lasaga, A. C. Crystal-growth kinetics of plagioclase in igneous systems: Isothermal H₂O-saturated experiments and extension of a growth model to complex silicate melts. *Am. Mineral.* **73**, 982–992 (1988).
- Harford, C. L. & Sparks, R. S. J. Recent remobilisation of shallow-level intrusions on Montserrat revealed by hydrogen isotope composition of amphibole. *Earth Planet. Sci. Lett.* **185**, 285–297 (2001).
- Pallister, J. S., Hoblitt, R. P. & Reyes, A. G. A basalt trigger for the 1991 eruptions of Pinatubo volcano. *Nature* **356**, 426–428 (1992).
- Nakada, S. & Motomura, Y. Petrology of the 1991–1995 eruption at Unzen: effusion pulsation and groundmass crystallisation. *J. Volcanol. Geotherm. Res.* **89**, 173–196 (1999).
- Sato, H., Nakada, S., Fujii, T., Nakamura, M. & Suzuki-Kamata, K. Groundmass pargasite in the 1991–1995 dacite of Unzen volcano: phase stability experiments and volcanological implications. *J. Volcanol. Geotherm. Res.* **89**, 197–212 (1999).
- Nekvasil, H. Ascent of felsic magmas and formation of rapakivi. *Am. Mineral.* **76**, 1279–1290 (1991).
- Hibbard, M. J. The magma mixing origin of mantle feldspars. *Contrib. Mineral. Petrol.* **76**, 158–170 (1981).
- Stimac, J. A. & Wark, D. A. Plagioclase mantles on sanidine in silicic lavas, Clear Lake, California: Implications for the origin of rapakivi texture. *Geol. Soc. Am. Bull.* **104**, 728–744 (1992).
- D'Lemos, R. S. Mixing between granitic and dioritic crystal mushes, Guernsey, Channel Islands, UK. *Lithos* **38**, 233–257 (1996).

- Vernon, R. H. Microgranitoid enclaves in granites—globules of hybrid magma quenched in a plutonic environment. *Nature* **309**, 438–439 (1983).
- Furman, T. & Spera, F. J. Co-mingling of acid and basic magma with implications for the origin of mafic I-type xenoliths: field and petrochemical relations of an unusual dike complex at Eagle Lake, Sequoia National Park, California, USA. *J. Volcanol. Geotherm. Res.* **24**, 151–178 (1985).
- Blundy, J. & Sparks, R. S. J. Petrogenesis of mafic inclusions in granitoids of the Adamello Massif. *J. Petrol.* **33**, 1039–1104 (1992).

Acknowledgements

S.C. acknowledges a NERC studentship, R.S.I.S. a NERC professorship and M.R.C. support from the Gruppo Nazionale de Vulcanologia.

Correspondence and requests for materials should be addressed to S.C. (e-mail: Susan.Couch@bristol.ac.uk).

Anaerobic benzene oxidation coupled to nitrate reduction in pure culture by two strains of *Dechloromonas*

John D. Coates, Romy Chakraborty, Joseph G. Lack, Susan M. O'Connor, Kimberly A. Cole, Kelly S. Bender & Laurie A. Achenbach

Department of Microbiology, Southern Illinois University, Carbondale, Illinois 62901, USA

Benzene contamination is a significant problem. It is used in a wide range of manufacturing processes and is a primary component of petroleum-based fuels. Benzene is a hydrocarbon that is soluble, mobile, toxic and stable, especially in ground and surface waters. It is poorly biodegraded in the absence of oxygen. However, anaerobic benzene biodegradation has been documented under various conditions. Although benzene biomineralization has been demonstrated with nitrate¹, Fe(III)^{2–5}, sulphate^{6,7} or CO₂^{8,9} as alternative electron acceptors, these studies were based on sediments or microbial enrichments. Until now there were no organisms in pure culture that degraded benzene anaerobically. Here we report two *Dechloromonas* strains, RCB and JJ, that can completely mineralize various mono-aromatic compounds including benzene to CO₂ in the absence of O₂ with nitrate as the electron acceptor. This is the first example, to our knowledge, of an organism of any type that can oxidize benzene anaerobically, and we demonstrate the potential applicability of these organisms to the treatment of contaminated environments.

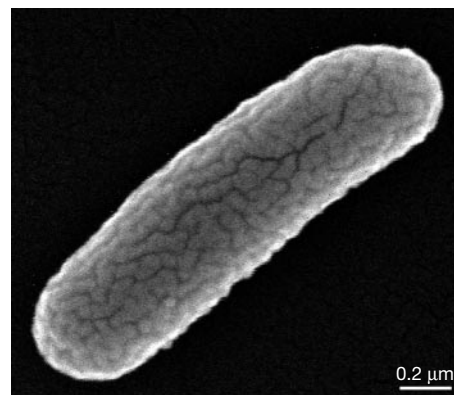


Figure 1 Scanning electron micrograph of anaerobically grown cells of strain RCB.

Benzene is one of the most prevalent organic contaminants in groundwaters¹⁰ and is of major concern owing to its toxicity and relatively high solubility. Numerous benzene-degrading aerobic microorganisms have been identified, the most notable of which are the *Pseudomonas* species, which may account for up to 87% of the petrol-degrading microorganisms in contaminated aquifers¹¹. However, soils and sediments contaminated with benzene frequently develop extensive anaerobic zones^{10,12,13}. As a result, much attention has focused on anaerobic benzene degradation. In the last decade, studies have demonstrated that the hydrocarbon-degrading capacity of anaerobes is far greater than previously assumed^{2,6,14–16}. Pure cultures of anaerobic organisms that can degrade hydrocarbons such as toluene, hexadecane and naphthalene have now been described^{17–20}. However, organisms capable of anaerobic benzene degradation have been elusive and this metabolism was observed only in sediment studies^{2–6,9,15,16} or with microbial enrichments^{1,7,8,21}. Organisms capable of anaerobic benzene degradation had not been isolated. Furthermore, even though anaerobic benzene degradation has been demonstrated under various conditions in the absence of O₂ (refs 1–3, 8, 9, 15, 16), no specific organisms or genera had been associated with the metabolism. Although Fe(III)-reducing sediments that degrade benzene^{4,5} are enriched with organisms of the family Geobacteraceae²², there is no direct evidence to show that members of this family are capable of benzene degradation.

Here we report on the isolation and characterization of *Dechloromonas* strains RCB and JJ, which oxidize benzene to CO₂ in the absence of O₂. These organisms were enriched and isolated from two diverse environments on the basis of very different metabolic abilities.

Strain RCB was enriched as a hydrocarbon-oxidizing chlorate reducer. Enrichments were initiated at 30 °C in anoxic artificial medium²³ with 4-chlorobenzoate (0.5 mM) as the electron donor and chlorate (10 mM) as the electron acceptor. Inocula were taken from sediments collected from the Potomac River, Maryland, USA, which contains an active microbial population that can reduce chlorate or perchlorate ((per)chlorate)^{24,25} and an anaerobic hydrocarbon-degrading population (J.D.C., unpublished observations). Positive enrichments were identified by microscopic observation of increasing cell density over time. After 8 d, cell numbers increased by at least one log unit. The cultures were transferred (10% v/v) into freshly prepared media. After several transfers in chlorobenzoate–chlorate medium, the enrichments were inoculated into an anoxic agar dilution series^{23,24} of the same medium amended with 2% noble agar. Colonies of strain RCB were isolated after six weeks' incubation.

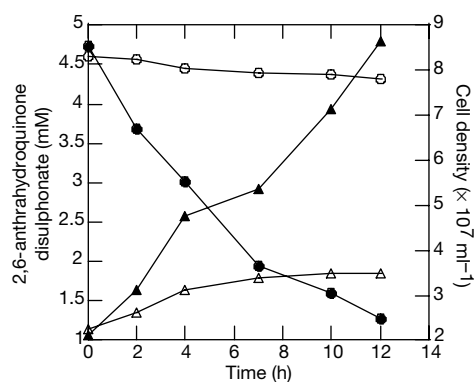


Figure 2 Growth of strain RCB (mean of triplicate determinations) with AHDS (5 mM) as the electron donor and chlorate (10 mM) as the electron acceptor. Acetate (0.1 mM) was added as a suitable carbon source. Filled circles, AHDS concentration in growth culture; open circles, AHDS concentration in heat-killed controls; filled triangles, cell density in growth culture; open triangles, cell density in heat-killed controls.

Strain JJ was isolated as a humic-substances-oxidizing nitrate reducer. Enrichments were initiated at 30 °C in anoxic artificial medium²³ with 2,6-anthrahydroquinone disulphonate (AHDS) (5 mM) as the electron donor and nitrate (5 mM) as the electron acceptor with 0.1 mM acetate as a carbon source. AHDS medium was prepared from anoxic freshwater medium containing 2,6-anthraquinone disulphonate reduced with palladium-coated aluminium chips and H₂ (ref. 26). Inocula were taken from sediments collected from Campus Lake, Southern Illinois University. Positive enrichments were identified by colour change in the media (red to tan) as the AHDS was oxidized. The cultures were transferred (10% v/v) into freshly prepared anoxic AHDS–nitrate medium twice over three weeks and then inoculated into an anoxic agar dilution series^{23,24} of the same medium amended with 2% noble agar. Colonies of strain JJ were isolated after three weeks' incubation.

Cells of strains RCB and JJ are gram-negative rods 1.8 × 0.5 μm (Fig. 1) that are facultatively anaerobic, completely oxidizing and non-fermenting. Both strains grew with air or nitrate as electron acceptors. Nitrate was reduced to N₂ gas. Strain RCB also grew with chlorate or perchlorate, completely reducing them to chloride. Both strains coupled growth to the complete oxidation of various simple organic acids with nitrate as the sole electron acceptor, including formate (10 mM), acetate (10 mM), propionate (5 mM), butyrate (5 mM), lactate (10 mM), succinate (5 mM), pyruvate (10 mM) and benzoate (0.5 mM). Both strains also oxidized toluene (1 mM). Both strains grew anaerobically on nitrate with the reduced analogue for humic substances, AHDS, as the electron donor and acetate (0.1 mM) as the carbon source (Fig. 2). Only slight growth was observed if the AHDS was omitted, which was probably due to the presence of the small amount of acetate (0.1 mM) added as a carbon source (Fig. 2). No AHDS oxidation occurred in the absence of an electron acceptor.

Phylogenetically, strains RCB and JJ are closely related to each other (98.1% 16S ribosomal DNA sequence similarity) and both are members of the *Dechloromonas* subgroup in the β subclass of the Proteobacteria²⁷. Their closest relative is *Ferribacterium limneticum* (Fig. 3), which is an anomaly in this subgroup²⁸ as it is a strictly anaerobic Fe(III) reducer. All other members of the subgroup are facultatively anaerobic (per)chlorate reducers^{27,28}.

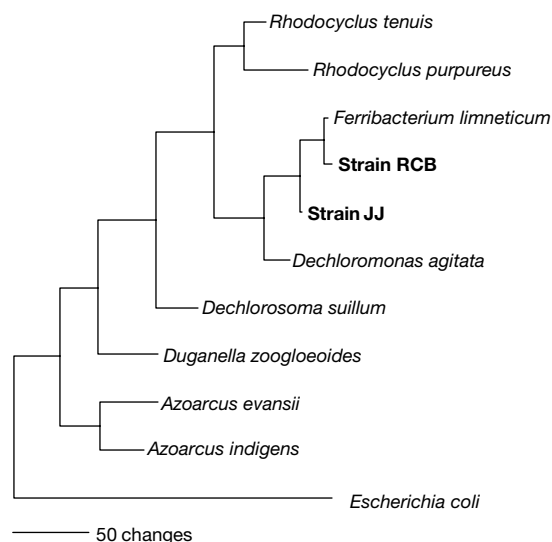


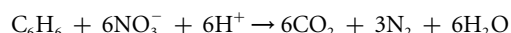
Figure 3 Phylogenetic tree of the 16S rDNA sequences of strains RCB and JJ and their closest relatives resulting from a heuristic search with parsimony analysis. The same topology was obtained from distance and from maximum likelihood, and was supported by bootstrap analysis.

Both strains RCB and JJ oxidized benzene anaerobically with nitrate as the electron acceptor (Fig. 4a). ^{14}C -labelled benzene was oxidized to $^{14}\text{CO}_2$. No ^{14}C benzene oxidation occurred if the electron acceptor was omitted (Fig. 4a). In the case of strain RCB, once $^{14}\text{CO}_2$ production ceased, 45% of the original ^{14}C label could be accounted for as $^{14}\text{CO}_2$. Analysis by high-performance liquid

chromatography (HPLC) indicated that 47% of the initial benzene remained undegraded in the culture medium, confirmed by comparison with heat-killed controls. This result accounts for 92% of the initial ^{14}C label, showing that benzene is completely oxidized to CO_2 . In support, HPLC analysis coupled to liquid scintillation counting indicated that the residual ^{14}C label in the aqueous phase was present only as benzene.

In parallel incubations with the aerobic benzene-oxidizer *Pseudomonas* strain JS-150, $^{14}\text{CO}_2$ production was not observed unless O_2 was added (data not shown), demonstrating that benzene oxidation by strains RCB and JJ was anaerobic and not the result of O_2 contamination. Because of the similar morphology, phylogeny and physiology of these two strains, strain RCB was selected for a more detailed characterization of anaerobic benzene oxidation coupled to nitrate reduction.

With nitrate as the sole electron acceptor, the oxidation of $163 \pm 19 \mu\text{M}$ benzene ($\pm$ standard deviation, $n = 3$) by strain RCB resulted in the reduction of $843 \pm 64 \mu\text{M}$ nitrate ($\pm$ standard deviation, $n = 3$). This represents greater than 86% of the theoretical ratio for benzene oxidation coupled to nitrate reduction according to:



In an active benzene-degrading culture of strain RCB with nitrate as the electron acceptor, increase in cell number was concomitant with benzene disappearance (Fig. 4b). Minimal growth occurred in the absence of benzene (Fig. 4b) and was probably the result of carry-over from the inoculum. The small amount of benzene lost in the heat-killed controls was due to absorption into the butyl-rubber stoppers^{8,16}. When the active culture from the growth curve shown in Fig. 4b was used to inoculate fresh medium, benzene was again rapidly degraded and cell increase was concomitant with benzene removal (Fig. 4c), demonstrating that strain RCB could grow and be transferred in media with benzene as the sole electron donor and nitrate as the sole electron acceptor. When strain RCB was grown anaerobically with nitrate and universally labelled (UL)- ^{14}C benzene, 2% of the ^{14}C label was associated with the retentate after the culture was filtered through a filter with a pore size of $0.2 \mu\text{m}$. A similar filtration of a heat-killed control retained 0.8% of the ^{14}C label, suggesting that 1.2% of the ^{14}C label was incorporated into biomass. In contrast, 3.0% of the ^{14}C label was incorporated into the biomass in an aerobically grown culture. These low values of incorporation of ^{14}C label into biomass are similar to values reported previously for a nitrate-reducing, benzene-oxidizing

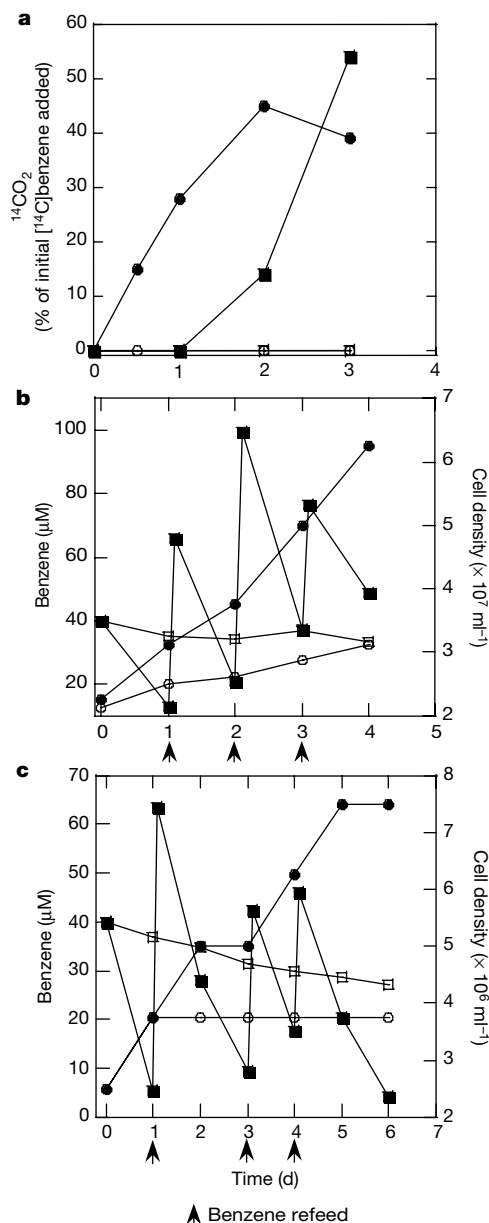


Figure 4 Anaerobic oxidation of benzene by strains JJ and RCB (mean of triplicate determinations). **a**, ^{14}C benzene is completely oxidized to $^{14}\text{CO}_2$ by strains RCB (filled circles) and JJ (filled squares) in the absence of oxygen with nitrate as the electron acceptor. No $^{14}\text{CO}_2$ production was observed with either strain (RCB, open circles; JJ, open squares) in the absence of nitrate or if the cells were heat-killed. **b**, Anaerobic growth and benzene oxidation by strain RCB with nitrate as the electron acceptor (filled squares, benzene concentration in the growth culture; open squares, benzene concentration in the heat-killed controls; filled circles, cell numbers in the growth culture; open circles, cell numbers in the absence of benzene). **c**, Growth and benzene oxidation in fresh anoxic culture media with nitrate as the electron acceptor after inoculation (10% v/v) from the active growth culture tubes from **b** (filled squares, benzene concentration in the growth culture; open squares, benzene concentration in the heat-killed controls; filled circles, cell numbers in the growth culture; open circles, cell numbers in the absence of benzene).

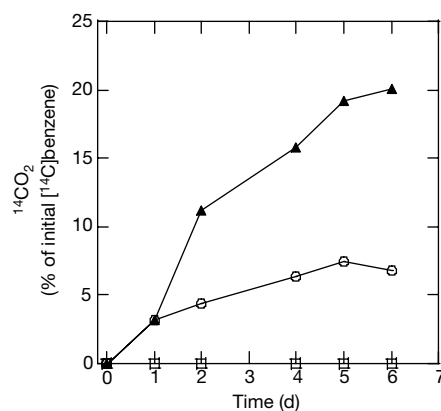


Figure 5 Degradation of ^{14}C benzene in anoxic sediments amended with nitrate (10 mM) in the presence and absence of strain RCB (mean of triplicate determinations). Filled triangles, sediment amended with both strain RCB and nitrate. Open circles, nitrate-amended sediment without strain RCB; open squares, unamended sediment without strain RCB.

enrichment culture (5–8%)¹ and are supportive of the low cell yield for anaerobic growth on benzene observed for strain RCB (Fig. 4b).

In order to determine the ability of these organisms to attenuate benzene in the natural environment, anoxic aquatic sediments amended with [¹⁴C]benzene and nitrate were inoculated with an active culture of strain RCB. Geochemical analysis of the sediment revealed that the indigenous nitrate and sulphate concentrations were 22 and 867 μM, respectively. No chlorate was present. The total iron content was 38 mM, and 45% of this (= 17 mM) was present as Fe(II). Determination of the terminal electron-accepting process in these sediments with [¹⁴C]acetate¹⁶ indicated that Fe(III) reduction was predominant (data not shown). Samples amended with both strain RCB and nitrate showed rapid evolution of ¹⁴CO₂ from UL-[¹⁴C]benzene (Fig. 5). Benzene oxidation also occurred in the unamended sediments (Fig. 5). This was unlikely to be due to O₂ contamination because the high concentrations of Fe(II) in the sediment would rapidly remove O₂. The indigenous benzene oxidation was inhibited by the addition of nitrate when strain RCB was omitted (Fig. 5). Similar inhibition by nitrate addition has previously been reported for a benzene-oxidizing, sulphate-reducing sediment²⁹.

Anaerobic benzene oxidation by microbes has been extensively documented in the last decade; however, until now it had only been observed in sediment studies or with enrichment cultures^{1–9,21}. Here we describe two isolates that can couple the anaerobic oxidation of benzene to the reduction of nitrate. Little is known about either microbial oxidation of humic substances or microbial (per)chlorate reduction. Previous studies in our lab have demonstrated that the phylogeny of microorganisms capable of these types of metabolism is diverse^{23,24,26}. Very little is known of their other metabolic capabilities. Both strains JJ and RCB are members of the *Dechloromonas* genus even though these organisms were isolated from two different environments on the basis of very different metabolic capabilities (that is, microbial (per)chlorate reduction and microbial humic-substances oxidation). Previous studies by our group have demonstrated the ubiquity of the *Dechloromonas* species, and members of this genus have been identified in a broad range of environments including pristine and contaminated soils and sediments²⁴, and even in samples collected from Antarctica (L.A.A., unpublished observations). The results presented here indicate that the members of this genus offer great potential not only for the attenuation of perchlorate but also for the treatment of benzene in contaminated environments. □

Methods

Medium preparation

All media and solutions were prepared using strict anaerobic techniques as described^{6,23,24}. All culturing was done in sealed serum vials with N₂/CO₂ (80/20, v/v) in the headspace. Basal medium used for the isolation of strain RCB was amended from a sterile aqueous anoxic stock solution of 4-chlorobenzoate (100 mM) to give a final concentration of 0.5 mM when appropriate. Basal medium used for the isolation of strain JJ was prepared from anoxic freshwater medium containing 2,6-anthraquinone disulphonate reduced with palladium-coated aluminium chips and H₂, as described²⁶.

In general, alternative electron donors tested were added from sterile aqueous anoxic stock solutions. When necessary, benzene was added from an anoxic aqueous stock solution prepared as described⁶. Growth was determined by observation of cell density increase and was confirmed by direct cell counts under oil-immersion phase-contrast microscopy. Results were compared with those of unamended controls. [¹⁴C]Benzene (1 μCi) was added from an anoxic stock solution as described^{25,30}.

Sediment experiments

Fresh sediments were collected from Campus Lake, Southern Illinois University, and stored under N₂ at 4 °C until used. Analyses of geochemical and terminal electron-accepting processes were performed directly before use as described^{14–16}. All experiments were performed in triplicate with strict anaerobic techniques in sealed 35-ml serum vials amended with 5 g sediment under an N₂ headspace. The inoculum size was 10% v/w of anaerobically grown cells with nitrate (10 mM) as the sole electron acceptor. [¹⁴C]benzene (1 μCi) was added from a stock solution⁶ and benzene concentrations were determined by gas chromatography with gas proportional counting detection (GC-GPC)^{25,30}. Nitrate was amended from an anoxic sterile stock solution.

Electron microscopy

Scanning electron micrographs were prepared from cells grown anaerobically with acetate (10 mM) as the electron donor and nitrate (10 mM) as the electron acceptor as described²⁵ and viewed with a Hitachi S570 scanning electron microscope at 20 kV.

Analytical techniques

Concentrations of ¹⁴CO₂ in the headspace of cultures amended with [¹⁴C]benzene were determined by GC-GPC^{25,30}. Benzene concentrations in growth cultures were determined by gas chromatography of 0.1-ml headspace samples as described⁶. Concentrations of aqueous-phase benzene and potential catabolic intermediates were determined by HPLC with an LC-18 Supelcosil column with ultraviolet detection at 254 nm. The mobile phase was 60/40 (v/v) methanol/water. If [¹⁴C]benzene was used, eluted peaks were collected directly into 5 ml of liquid scintillant and analysed for ¹⁴C label by liquid scintillation counting with a Beckman LS 6000IC. Nitrate concentrations were determined by ion chromatography of aqueous samples with a Dionex DX500 equipped with an AS9-SC column using a mobile phase of sodium carbonate (2 mM) and sodium bicarbonate (7.5 mM) at a flow rate of 2 ml min⁻¹. AHDS concentrations were determined by colorimetric assay at 450 nm as described²⁶. Cell numbers were determined by direct microscopic counting under oil immersion.

Phylogenetic analysis

Polymerase chain reaction, sequencing and analysis of the 16S ribosomal RNA genes was performed as described^{24,27}. Sequence entry and manipulation was performed with the MacVector 6.5 sequence analysis software program (Oxford Molecular Group). Sequences of select 16S rRNAs were downloaded from the Ribosomal Database Project and GenBank into the computer program SeqApp. 16S rDNA sequences of strains RCB and JJ were manually added to the alignment using secondary structure information for proper alignment (alignment available from J.D.C.). Only those regions sequenced in all of the organisms (1,456 nucleotides) were used in the subsequent phylogenetic analyses. Distance, parsimony and maximum likelihood analysis of the aligned sequences was performed with PAUP* 4.0. Bootstrap analysis was conducted on 100 replications using a heuristic search strategy to assess the confidence level of various clades. GenBank accession numbers for sequences represented in Fig. 3 are: *Rhodocyclus tenuis*, D16209; *R. purpureus*, M34132; *Ferribacterium limneticum*, Y17060; *Dechloromonas agitata*, AF047462; *Dechlorosoma suillum*, AF170348; *Duganella zoogloeoides* (previously *Zoogloea ramigera*), X74913; *Azoarcus evansii*, X77679; *A. indigens*, L15531; and *Escherichia coli*, J01859).

Received 6 February; accepted 1 May 2001.

- Burland, S. M. & Edwards, E. A. Anaerobic benzene biodegradation linked to nitrate reduction. *Appl. Environ. Microbiol.* **65**, 529–533 (1999).
- Lovley, D. R., Woodward, J. C. & Chapelle, F. H. Stimulated anoxic biodegradation of aromatic hydrocarbons using Fe(III) ligands. *Nature* **370**, 128–131 (1994).
- Lovley, D. R., Woodward, J. C. & Chapelle, F. H. Rapid anaerobic benzene degradation with a variety of chelated Fe(III) forms. *Appl. Environ. Microbiol.* **62**, 288–291 (1996).
- Anderson, R. T. & Lovley, D. R. Naphthalene and benzene degradation under Fe(III)-reducing conditions in petroleum-contaminated aquifers. *Bioremed. J.* **3**, 121–135 (1999).
- Anderson, R. T., Rooney-Varga, J., Gaw, C. V. & Lovley, D. R. Anaerobic benzene oxidation in the Fe(III) reduction zone of petroleum-contaminated aquifers. *Environ. Sci. Technol.* **32**, 1222–1229 (1998).
- Lovley, D. R., Coates, J. D., Woodward, J. C. & Phillips, E. J. P. Benzene oxidation coupled to sulfate reduction. *Appl. Environ. Microbiol.* **61**, 953–958 (1995).
- Weiner, J. & Lovley, D. R. Anaerobic benzene degradation in petroleum-contaminated sediments after inoculation with a benzene-oxidizing enrichment. *Appl. Environ. Microbiol.* **64**, 775–778 (1998).
- Grbic-Galic, D. & Vogel, T. Transformation of toluene and benzene by mixed methanogenic cultures. *Appl. Environ. Microbiol.* **53**, 254–260 (1987).
- Weiner, J. & Lovley, D. R. Rapid benzene degradation in methanogenic sediments from a petroleum-contaminated aquifer. *Appl. Environ. Microbiol.* **64**, 1937–1939 (1998).
- Anderson, R. T. & Lovley, D. R. Ecology and biogeochemistry of *in situ* groundwater bioremediation. *Adv. Microb. Ecol.* **15**, 289–350 (1997).
- Ridgeway, H. F., Safarik, J., Phipps, D., Carl, P. & Clark, D. Identification and catabolic activity of well-derived gasoline-degrading bacteria and a contaminated aquifer. *Appl. Environ. Microbiol.* **56**, 3565–3575 (1990).
- Christensen, T., Kjeldsen, P., Albrechtsen, H. & Heron, G. Attenuation of pollutants in landfill leachate polluted aquifers. *Crit. Rev. Environ. Sci. Technol.* **24**, 119–202 (1994).
- Lovley, D. R. Potential for anaerobic bioremediation of BTEX in petroleum-contaminated aquifers. *J. Ind. Microbiol.* **18**, 75–81 (1997).
- Coates, J. D., Anderson, R. T. & Lovley, D. R. Anaerobic oxidation of polycyclic aromatic hydrocarbons under sulfate-reducing conditions. *Appl. Environ. Microbiol.* **62**, 1099–1101 (1996).
- Coates, J. D., Woodward, J., Allen, J., Philp, P. & Lovley, D. R. Anaerobic degradation of polycyclic aromatic hydrocarbons and alkanes in petroleum-contaminated marine harbor sediments. *Appl. Environ. Microbiol.* **63**, 3589–3593 (1997).
- Coates, J. D., Anderson, R. T. & Lovley, D. R. Anaerobic hydrocarbon degradation in petroleum-contaminated harbor sediments under sulfate-reducing and artificially imposed iron-reducing conditions. *Environ. Sci. Technol.* **30**, 2784–2789 (1996).
- Galushko, A., Minz, D., Schink, B. & Widdel, F. Anaerobic degradation of naphthalene by a pure culture of a novel type of marine sulfate-reducing bacterium. *Environ. Microbiol.* **1**, 415–420 (1999).
- Rabus, R., Nordhaus, R., Ludwig, W. & Widdel, F. Complete oxidation of toluene under strictly anoxic conditions by a new sulfate-reducing bacterium. *Appl. Environ. Microbiol.* **59**, 1444–1451 (1993).
- Lovley, D. R. & Lonergan, D. J. Anaerobic oxidation of toluene, phenol, and p-cresol by the dissimilatory iron-reducing organism, GS-15. *Appl. Environ. Microbiol.* **56**, 1858–1864 (1990).

20. Coates, J. D., Phillips, E. J. P., Lonergan, D. J., Jenter, H. & Lovley, D. R. Isolation of *Geobacter* species from a variety of sedimentary environments. *Appl. Environ. Microbiol.* **62**, 1531–1536 (1996).
21. Vogel, T. M. & Grbic-Galic, D. Incorporation of oxygen from water into toluene and benzene during anaerobic fermentative transformation. *Appl. Environ. Microbiol.* **52**, 200–202 (1986).
22. Rooney-Varga, J. N., Anderson, R. T. & Fraga, J. L. Microbial communities associated with anaerobic benzene degradation in a petroleum-contaminated aquifer. *Appl. Environ. Microbiol.* **65**, 3056–3063 (1999).
23. Bruce, R. A., Achenbach, L. A. & Coates, J. D. Reduction of (per)chlorate by a novel organism isolated from a paper mill waste. *Environ. Microbiol.* **1**, 319–331 (1999).
24. Coates, J. D. *et al.* The ubiquity and diversity of dissimilatory (per)chlorate-reducing bacteria. *Appl. Environ. Microbiol.* **65**, 5234–5241 (1999).
25. Coates, J. D., Bruce, R. A., Patrick, J. A. & Achenbach, L. A. Hydrocarbon bioremediative potential of (per)chlorate-reducing bacteria. *Bioremed. J.* **3**, 323–334 (1999).
26. Lovley, D. R., Fraga, J. L., Coates, J. D. & Blunt-Harris, E. L. Humics as an electron donor for anaerobic respiration. *Environ. Microbiol.* **1**, 89–98 (1999).
27. Achenbach, L. A., Bruce, R. A., Michaelidou, U. & Coates, J. D. *Dechloromonas agitata* N. N. gen., sp. nov. and *Dechlorosoma suillum* N. N. gen., sp. nov. two novel environmentally dominant (per)-chlorate-reducing bacteria and their phylogenetic position. *Int. J. Syst. Evol. Microbiol.* **51**, 527–533 (2001).
28. Achenbach, L. A. & Coates, J. D. Disparity between bacterial phylogeny and physiology. *ASM News* **66**, 714–716 (2000).
29. Anderson, R. T. & Lovley, D. R. Anaerobic bioremediation of benzene under sulfate-reducing conditions in a petroleum contaminated aquifer. *Environ. Sci. Technol.* **34**, 2261–2266 (2000).
30. Coates, J. D., Bruce, R. A. & Haddock, J. D. Anoxic bioremediation of hydrocarbons. *Nature* **396**, 730 (1998).

Acknowledgements

This work was supported in part by funding to J.D.C. by the US Office of Naval Research. Funding to J.D.C. and L.A.A. was also from the US Department of Defense SERDP Program.

Correspondence and requests for materials should be addressed to J.D.C. (e-mail: jcoates@micro.siu.edu). GenBank accession numbers for strains RCB and JJ are AY032610 and AY032611, respectively.

Life-history traits of voles in a fluctuating population respond to the immediate environment

Torbjørn Ergon*, Xavier Lambin† & Nils Chr. Stenseth*

* Division of Zoology, Department of Biology, University of Oslo, PO Box 1050 Blindern, 0316 Oslo, Norway

† Department of Zoology, University of Aberdeen, Tillydrone Avenue, Aberdeen AB24 2TZ, UK

Life-history traits relating to growth and reproduction vary greatly among species and populations^{1,2} and among individuals within populations³. In vole populations, body size and age at maturation may vary considerably among locations and among years within the same location^{4–8}. Individuals in increasing populations are typically larger and start reproduction earlier in the spring than those in declining populations^{6–8}. The cause of such life-history variation within populations has been subject of much discussion^{7,9,10}. Much of the controversy concerns whether the memory of past conditions, leading to delayed effects on life-history traits, resides in the environment (for example, predators^{11,12}, pathogens¹³ or food^{14,15}) or intrinsically within populations or individuals (age distribution^{16,17}, physiological state³, genetic¹⁸ or maternal effects^{19,20}). Here we report from an extensive field transplant experiment in which voles were moved before the breeding season between sites that differed in average overwintering body mass. Transplanted voles did not retain the characteristics of their source population, and we demonstrate an over-riding role of the immediate environment in shaping life-history traits of small rodents.

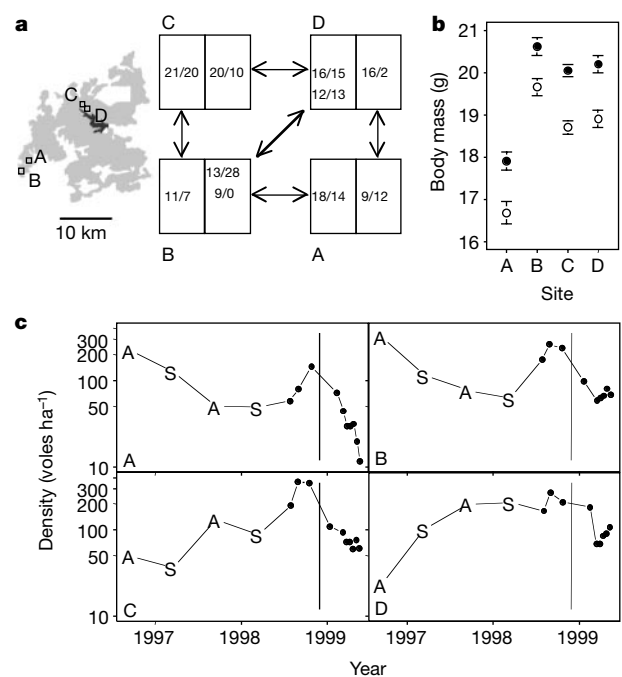


Figure 1 Design of the study. **a**, Groups of voles were swapped between half-sides of four 1-ha unfenced trapping grids in early winter (November/December). Arrows, pairs of transplanted groups. A second transplant was done between two of the sites (B and D) in February because few voles settled in some groups. Numbers in the boxes show number of males/females that were successfully transplanted between grids and captured at least once during the spring. **b**, Average body mass ($\pm$ s.e.) at the time of transplant in November/December among animals that had not previously bred. Open circles, females; filled circles, males. **c**, Density trajectories at the four sites plotted on a log₁₀ scale. Density estimates denoted by S (spring) and A (autumn) are based on vole sign indices²¹. Estimates plotted as filled circles are obtained by the robust design model²⁷ of capture–mark–recapture data. Vertical lines show the time of transplant. Density trajectories in neighbouring control areas showed similar patterns (see Supplementary Information).

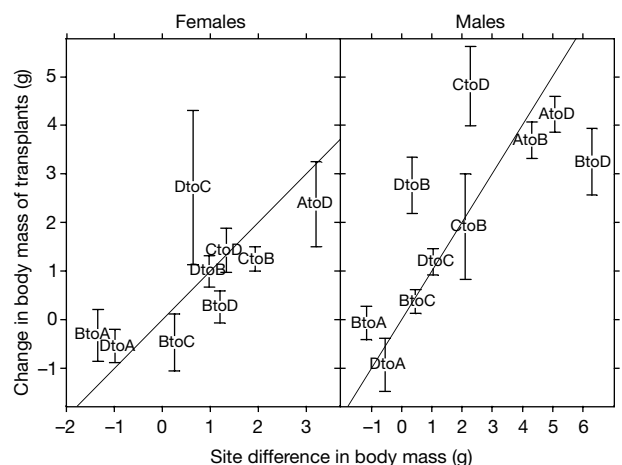


Figure 2 Average ($\pm$ s.e.) individual change in body mass of transplanted voles recaptured in January/February during the first session after transplanting plotted against difference in average body mass at the sites. The site differences in body mass are calculated as average body mass in the target population during the first trapping session after the transplant minus average body mass of transplants at capture during the transplantation. Voles in the transplant group are excluded from the target population when calculating the site differences. Hence, measurements are independent. Line is the unity slope through the origin, indicating perfect convergence. Plotted labels denote ‘source-to-target’ transplant groups. See Supplementary Information.

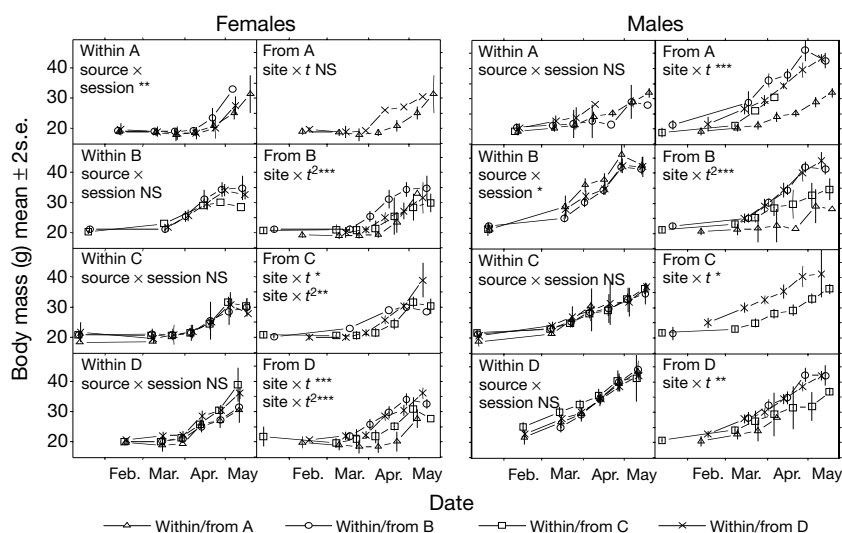


Figure 3 Average body mass (± 2 s.e.) over the spring after the transplant. Each line represents a group of animals that were living in the same site and were coming from the same site before transplant. Left column, time trends are grouped according to the site the animals were living in; right column, the same time trends are grouped according to site of origin. The data were analysed with a repeated-measures analysis with individuals as random subjects (SAS Proc MIXED²⁸). An unstructured covariance matrix (UN(3)²⁸) was selected by the lowest AIC. When studying the 'source' effect, a continuous time variable (t , number of days since the first trapping session) was used because sites were not trapped on the same dates. To account for the exponential pattern in the growth curves,

effects of both time (t) and t^2 were included. This gave homogenous residuals that appeared normally distributed. A categorical time variable (trapping session) was used when growth patterns within sites were compared. Because there are missing combinations of 'site' and 'source', the population effects were examined with separate analysis for each group of 'site' and 'source'. Asterisks represent significance levels of the interaction effect of population and time (triple asterisk; $P < 0.001$, double asterisk; $0.001 < P < 0.01$, single asterisk, $0.01 < P < 0.05$). NS, not selected by the lowest AIC. Effects of whether the voles had been transplanted or immigrated to the sites were not significant at the 10% level.

This field experiment was carried out in the Kielder Forest region, northern England, where fluctuating populations of field voles (*Microtus agrestis* L.) inhabit distinct 'clear-cuts' showing asynchronous dynamics within the region^{21,22}. We successfully transplanted 266 voles between four clear-cuts in the forest system (Fig. 1a). At the time of transplant, average body mass at the site with the heaviest voles was 18% greater than at the site with the smallest voles (Fig. 1b). Although an estimated 98% of residents were removed before the release of the transplanted voles, 281 non-transplanted voles (mostly immigrants) were recorded in the first trapping session 2–3 months after the transplant (in January and February). A further 287 overwintering voles entered the trapping grids at some time during the spring.

We used capture–recapture data collected at two-week intervals throughout the spring to investigate differences in patterns of body mass development and onset of maturation between the different groups of voles. Trapping two to three months after the transplant showed that individual body mass during midwinter changed towards the mean body mass at the site to which the voles were moved (Fig. 2). Voles in the sites with the highest midwinter body mass also grew faster and reached higher body mass by the end of the study than those in the sites with low midwinter body mass (Fig. 3). Individuals originating from different sites before the transplant but living in the same site showed no clear differences in body mass trajectories through the spring. However, there was large variation in individual growth rates between sites during the spring: voles originating from the same site but transplanted to different sites showed up to 60% higher body mass in one site than at another at the same dates. Hence, individuals adjusted their midwinter body size and spring growth rate according to their immediate environment.

The change in proportion of matured animals in different groups through the spring reveals a difference of around three weeks in onset of spring reproduction between sites (Fig. 4). There was no significant effect of source population on timing of maturation, but

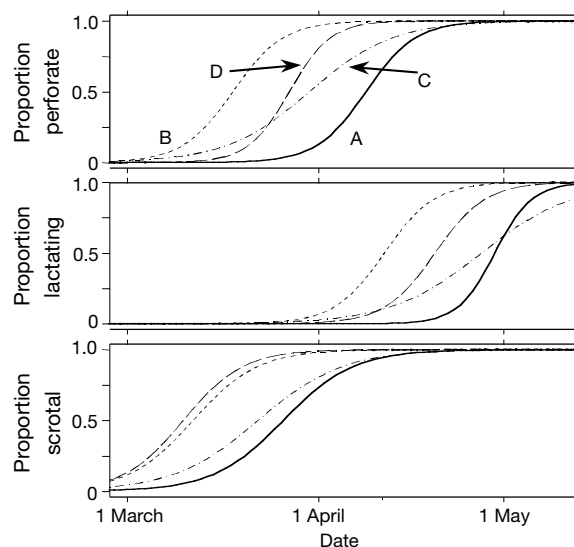


Figure 4 Onset of spring reproduction measured as proportions of perforate and lactating females (top and middle panels) and proportion of males with scrotal testes (bottom panel). The Figure shows fits from the logistic regression models (see Methods and Table 1). Effects of whether the voles were transplanted or immigrated and effects of source population were not significant (see Table 1) and hence not accounted for.

the site effect was highly significant (Table 1). There was a one-to-one correspondence of the sites ranked according to onset of spring reproduction and ranks of average overwinter body mass and spring growth. Reproduction started earlier in the sites with higher body mass. The onset of spring reproduction may contribute profoundly to the variation in population growth rate, and is generally

Table 1 Test statistics for logistic regression of proportion of matured voles in the trapping sessions

Sex (no. individuals)	Response	Factor	F	d.f.*	P	Deviance of model
Females (n = 121)	Proportion with perforate vagina	Site	11.31	3,113	1.5×10^{-6}	107.2
		Source	1.36	3,113	0.26	
	Proportion lactating	Site × date	2.76	3,107	0.046†	84.7
		Source × date	1.11	3,107	0.35	
Males (n = 145)	Proportion with scrotal testis	Site	10.03	3,137	5.1×10^{-6}	111.2
		Source	0.70	3,137	0.55	

All models include a 'date' effect. Models with either parallel or different slopes ('group × date' interactions) are chosen on the basis of the models' AIC²⁶. Because of imbalanced replication, 'site' is tested assuming an effect of 'source' and vice versa. Only voles that have been transplanted (that is, no immigrants) are included in the analysis.

*Degrees of freedom reflect number of individuals, not observations (see Methods).

†Main effects in a parallel slopes model: site, $F_{3,113} = 12.3$, $P = 5.0 \times 10^{-7}$; source, $F_{3,113} = 0.61$, $P = 0.6$; deviance, 101.0.

described as varying consistently between the phases in rodent fluctuations^{5,7}. Even though the variation in life-history traits documented here is correlated with the variation in population growth rate in the spring, there is no indication that this variation reflects present or past densities (Fig. 1c).

Our results show strong plasticity in life-history traits of voles. Individuals do not perform according to past environmental conditions. Instead, the environmental conditions during the winter and spring directly give rise to large differences in body mass and timing of maturation between sites. Thus, there is no need to invoke intrinsic mechanisms^{16,17,23} to explain the conspicuous variation in these life-history traits within vole populations. □

Methods

Transplanting voles

The transplant took place in November/December, when the voles were non-reproductive. Each 1-ha trapping grid was divided into two sides, and voles were swapped between half-sides of two different grids at a time (Fig. 1). Trapping continued for 5 d until an estimated 98% of the resident animals were captured. Voles were transferred between sites in wooden boxes containing hay, sawdust, carrots and whole wheat (two individuals in each box). After the last day of trapping, the boxes were opened but left in the field for at least one week. In total, 761 individually marked animals were swapped between the four locations. Of these, 266 individuals were recaptured during the following spring. Although voles that had not previously reproduced (91% of the transplanted voles) had a higher probability of being recaptured after transplant (odds ratio = 1.4, $\chi^2_1 = 4.84$, $P = 0.03$), reflecting similar effects on natural mortality, there was no significant difference in the effects of body mass and reproductive history on transplant success between the groups defined by 'target population', 'source population' and 'target × source group' (all interactions tested in logistic regression models selected by the lowest Akaike Information Criterion²⁶ (AIC); all $P > 0.1$). Thus, there is no reason to believe that differential transplant success will substantially bias our results. A detailed presentation on the analysis of transplant success is available as Supplementary Information.

Population monitoring

The populations at the four sites were monitored by capture-mark-recapture methods^{24,25}. At each site, 196 traps at 7-m intervals were pre-baited with carrots 2 d before the traps were set (baited with carrots and flaked barley). Food was removed from the traps after each session. The first trapping session after the transplant took place in January/February. Each site was thereafter visited every two weeks (except for the first interval; which varied in length between sites). In the autumn before the transplant, traps were checked every 6 h and not left set overnight. After the transplant, traps were set in the evening and checked at sunrise and sunset for the three consecutive sessions.

Comparing onset of reproduction

At each capture occasion, we recorded whether females had perforated vagina and whether they were lactating. Males were classified as mature when they had scrotal testes with visible *caput epididymis*. Effects on the timing of maturation were analysed by logistic regression of the proportion of matured animals on sampling date (recapture probability was ~0.9 and not significantly different between sites or source populations). As each individual will necessarily be recorded as matured on every subsequent capture after having reached maturity, the saturated model (zero deviance) is a model with individual specific intercepts and a common slope. Hence, we calculated the test statistic as $F = ((\text{Dev}_{H_0} - \text{Dev}_{H_A})/d.f._1)/((\text{Dev}_{H_A}/d.f._2))$ with d.f.₁ being the difference in number of parameters in the two models (H_0 and H_A) and d.f.₂ being the number of individuals minus the number of parameters in H_A , and compared this with the Fisher distribution with d.f.₁ and d.f.₂ degrees of freedom.

Received 12 December 2000; accepted 3 April 2001.

1. Stearns, S. C. *The Evolution of Life Histories* (Oxford Univ. Press, Oxford, 1992).
2. Roff, D. A. *The Evolution of Life Histories: Theory and Analysis* (Chapman & Hall, New York, 1992).

3. McNamara, J. M. & Houston, A. I. State-dependent life histories. *Nature* **380**, 215–221 (1996).
4. Agrell, J., Erlinge, S., Nelson, J. & Sandell, M. Body weight and population dynamics: cyclic demography in a noncyclic population of the field vole (*Microtus agrestis*). *Can. J. Zool.* **70**, 494–501 (1992).
5. Taitt, M. J. & Krebs, C. J. in *Biology of New World Microtus* (ed. Tamarin, R. H.) 567–620 (American Soc. Mammalogists Special Publications No. 8, Shippensburg, 1985).
6. Chitty, D. Mortality among voles (*Microtus agrestis*) at Lake Vyrnwy, Montgomeryshire, in 1936–9. *Phil. Trans. R. Soc. Lond. B* **236**, 505–552 (1952).
7. Krebs, C. J. & Myers, J. H. Population cycles in small mammals. *Adv. Ecol. Res.* **8**, 267–399 (1974).
8. Taitt, J. in *Winter Ecology of Small Mammals* (ed. Merritt, J. E.) 59–66 (Carnegie Museum of Natural History, Pittsburgh, 1984).
9. Krebs, C. J. Population cycles revisited. *J. Mamm.* **77**, 8–24 (1996).
10. Stenseth, N. C. Population cycles in voles and lemmings: density dependence and phase dependence in a stochastic world. *Oikos* **87**, 427–461 (1999).
11. Korpimäki, E. & Krebs, C. J. Predation and population cycles of small mammals. *BioScience* **46**, 754–764 (1996).
12. Mappes, T., Koskela, E. & Ylönen, H. Breeding suppression in voles under predation risk of small mustelids: laboratory or methodological artifact? *Oikos* **82**, 365–369 (1998).
13. Soveri, T. *et al.* Disease patterns in field and bank vole populations during a cyclic decline in central Finland. *Comp. Immunol. Microbiol. Infect. Dis.* **23**, 73–89 (2000).
14. Batzli, G. O. in *Wildlife 2001: Populations* (eds McCullough, D. R. & Barrett, R. H.) (Elsevier Applied Sciences, London, 1992).
15. Agrell, J., Erlinge, S., Nelson, J., Nilsson, C. & Persson, I. Delayed density-dependence in a small-rodent population. *Proc. R. Soc. Lond. B* **262**, 65–70 (1995).
16. Boonstra, R. Population cycles in microtines: the senescence hypothesis. *Evol. Ecol.* **8**, 196–219 (1994).
17. Tkadlec, E. & Zejda, J. Small rodent population fluctuations: The effects of age structure and seasonality. *Evol. Ecol.* **12**, 191–210 (1998).
18. Chitty, D. The natural selection of self-regulatory behaviour in animal populations. *Proc. Ecol. Soc. Aust.* **2**, 51–78 (1967).
19. Boonstra, R. & Boag, P. T. A test of the Chitty hypothesis: inheritance of life-history traits in meadow voles *Microtus pennsylvanicus*. *Evolution* **41**, 929–947 (1987).
20. Inchausti, P. & Ginzburg, L. R. Small mammal cycles in northern Europe: patterns and evidence for maternal effect hypothesis. *J. Anim. Ecol.* **67**, 180–194 (1998).
21. Lambin, X., Petty, S. J. & MacKinnon, J. L. Cyclic dynamics in field vole populations and generalist predation. *J. Anim. Ecol.* **69**, 106–118 (2000).
22. Lambin, X., Elston, D., Petty, S. & MacKinnon, J. Spatial patterns and periodic travelling waves in cyclic field vole, *Microtus agrestis*, populations. *Proc. R. Soc. Lond. B* **265**, 1491–1496 (1998).
23. Nelson, R. J. Photoperiod-nonresponsive morphs: a possible variable in microtine population-density fluctuations. *Am. Nat.* **130**, 350–369 (1987).
24. Nichols, J. D. Capture recapture models; using marked animals to study population dynamics. *BioScience* **42**, 94–102 (1992).
25. Lebreton, J.-D., Burnham, K. P., Clobert, J. & Anderson, D. R. Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies. *Ecol. Monogr.* **62**, 67–118 (1992).
26. Burnham, K. P. & Anderson, D. R. *Model Selection and Inference: a Practical Information-Theoretic Approach* (Springer, New York, 1998).
27. Kendall, W. L., Pollock, K. H. & Brownie, C. A likelihood-based approach to capture-recapture estimation of demographic parameters under the robust design. *Biometrics* **51**, 293–308 (1995).
28. Littell, R. C., Milliken, G. A., Stroup, W. W. & Wolfinger, R. D. *SAS Systems for Mixed Models* (SAS Institute, Cary, North Carolina, 1996).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

The project was funded by grants from the Norwegian Research Council to T.E. and by the Natural Environment Research Council (UK) to X.L. We thank the Forestry Commission for support; C. Griffin and J. Aars for assistance during the transplant; and R. Boonstra, L. Crespin, K. E. Hodges, R. A. Ims, R. Julliard, T. Klemola, P. Thompson, E. Tkadlec, H. Viljugrein and N. G. Yoccoz for constructive comments and advice.

Correspondence and requests for materials should be addressed to N.C.S. (e-mail: n.c.stenseth@bio.uio.no) and X.L. (e-mail: x.lambin@abdn.ac.uk).

Protein dispensability and rate of evolution

Aaron E. Hirsh & Hunter B. Fraser

Center for Computational Genetics and Biological Modeling,
Department of Biological Sciences, Stanford University, Stanford, California, USA
The authors contributed equally to this work

If protein evolution is due in large part to slightly deleterious amino acid substitutions^{1,2}, then the rate of evolution should be greater in proteins that contribute less to individual fitness. The rationale for this prediction is that relatively dispensable proteins should be subject to weaker purifying selection, and should therefore accumulate mildly deleterious substitutions more rapidly. Although this argument was presented³ over twenty years ago, and is fundamental to many applications of evolutionary theory⁴, the prediction has proved difficult to confirm. In fact, a recent study showed that essential mouse genes do not evolve more slowly than non-essential ones⁵. Thus, although a variety of factors influencing the rate of protein evolution have been supported by extensive sequence analysis^{6–12}, the relationship between protein dispensability and evolutionary rate has remained unconfirmed. Here we use the results from a highly parallel growth assay of single gene deletions in yeast¹³ to assess protein dispensability, which we relate to evolutionary rate estimates that are based on comparisons of sequences drawn from twenty-one fully annotated genomes. Our analysis reveals a highly significant relationship between protein dispensability and evolutionary rate, and explains why this relationship is not detectable by categorical comparison of essential versus non-essential proteins. The relationship is highly conserved, so that protein dispensability in yeast is also predictive of evolutionary rate in a nematode worm.

For estimates of protein dispensability, we drew upon the results from a high-throughput method for measuring the growth rates of yeast strains¹³. In this method, each strain contains a distinct single-gene deletion, tagged by a unique genetic barcode that is readable on a microarray. A large number of strains are grown in batch culture,

and the relative intensities of barcode signals are used to estimate the growth rates of strains. Growth rates have been made available for 562 homozygous single-gene deletion mutants. For each mutant, we estimated the deleted gene's 'fitness effect', f_i , as $1 - r_i/r_{\max}$, where r_i is the growth rate of the strain with gene i deleted, and $r_{\max}$ is the maximal growth rate present in the batch culture. Thus, fitness effect in principle ranges from 0, meaning a gene has no measurable effect, to 1, meaning the gene is essential. However, the barcode growth assay is not accurate when the fitness effect of a deletion is greater than 0.5 (ref. 13). Therefore, 14 values had to be dropped from the data set, leaving 548 estimated fitness effects ranging from 0 to 0.5.

To estimate the relative rates of evolution of yeast proteins, we relied on comparison between each *Saccharomyces cerevisiae* protein and an orthologous sequence in *Caenorhabditis elegans*. Such comparisons can be used to estimate the relative rates of evolution only if all sequence pairs share the same time since divergence; otherwise, divergence time is confounded with evolutionary rate. In order to ensure that the two members of each pair of proteins have been evolving separately since the fungi–animal split, and not since a gene-duplication that might have occurred before this split, we required sequence pairs to be reciprocal best hits (RBHs)¹⁴. All RBHs were isolated from the results of reciprocal searches for homologous sequences using Washington University Basic Local Alignment Search Tool for Proteins (WU-BLASTP)¹⁵, yielding a total of 1,531 yeast/worm sequence pairs. For each pair i we estimated an evolutionary distance, d_i , defined as the number of substitutions per amino acid site (see Methods). There were 119 yeast genes for which we had both a distance to a RBH (d_i) and a fitness effect of deletion (f_i). If the rate of evolution is slower in proteins that contribute more to individual fitness, d_i should decline as f_i increases. As shown in Fig. 1, a highly significant relationship is in fact observed (linear regression: $d_i = -3.482f_i + 2.806$, $P = 0.001$; Spearman rank correlation, $P = 2.6 \times 10^{-7}$).

Each evolutionary distance, d_i , comprises evolutionary change not only of yeast protein i , but also of the protein's RBH in the worm. By contrast, our estimates of protein fitness effects are derived from measurements made only on yeast. To obtain an estimate of evolutionary rate that depends only on the evolutionary lineage leading to yeast, we separated each distance, d_i , into two segments: one between yeast and the most recent common ancestor

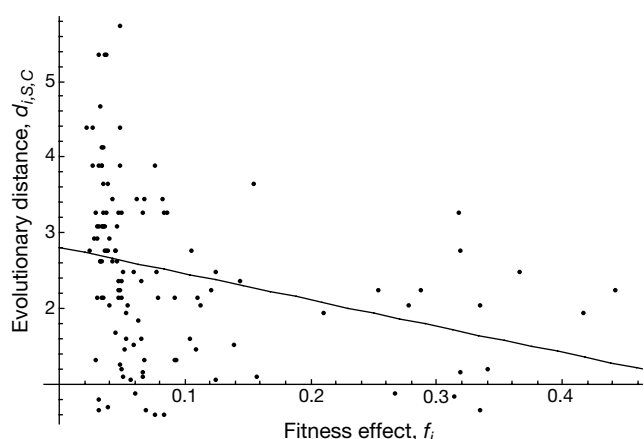


Figure 1 The relationship between a protein's contribution to fitness in yeast (f_i) and its rate of evolution as estimated by d_i , the evolutionary distance to the protein's reciprocal best hit in worm (linear regression: $d_i = -3.482f_i + 2.806$, $P = 0.001$; Spearman rank correlation, $P = 6.6 \times 10^{-5}$). The d_i values shown here were calculated by method 1, but the relationship was significant also when distances were calculated by method 2 (data not shown, $d_i = -18.422f_i - 40.663$, $P = 0.003$ in linear regression; Spearman rank correlation, $P = 6.6 \times 10^{-5}$).

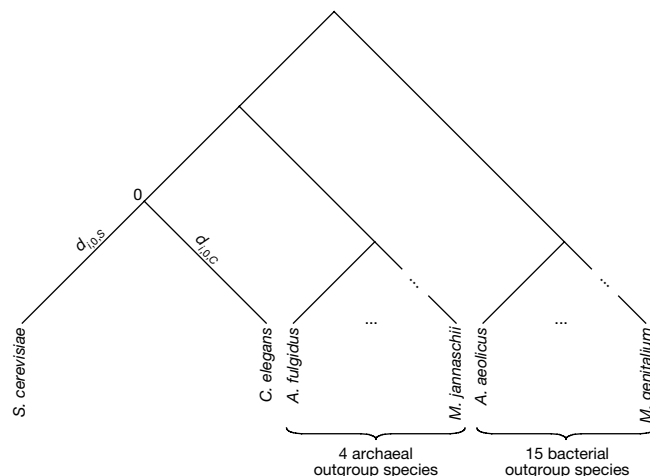


Figure 2 Schematic representation of phylogenetic relationships among yeast (*S. cerevisiae*), the worm (*C. elegans*), and outgroup species. Four archaeal and 15 bacterial species, representing 16 major phylogenetic lineages, were used as outgroups. See Methods for a complete list of genera. The distance from the fungi–animal ancestor (0) to yeast was estimated as $d_{i,S} = (d_{i,S,k} - d_{i,C,k} + d_{i,S,C})/2$, where k represents each of the 19 outgroups for which a probable orthologue of sequence i could be found.

of fungi and animals, and another between this ancestor and worm (Fig. 2). Each distance was decomposed by comparing yeast and worm sequences to outgroup species, a procedure that has been widely used in relative rate tests^{16–18}. Let $d_{i,j,k}$ denote the evolutionary distance between protein i in species j and its orthologue in species k . The yeast–worm distance, $d_{i,S,C}$, can be separated into two segments by using the relationship $d_{i,0,S} = (d_{i,S,k} - d_{i,C,k} + d_{i,S,C})/2$, where 0 represents the ancestor of fungi and animals, and k represents an outgroup species (Fig. 2). In order to increase the number of yeast–worm sequence pairs for which appropriate outgroup sequences could be found, and potentially to lower the variance of our estimate of $d_{i,0,S}$, we used outgroup sequences from 16 major lineages (see Methods). There were 48 yeast proteins for which a fitness effect, a worm RBH, and at least one suitable outgroup sequence were available. For each of these proteins, the point $(f_i, d_{i,0,S})$ is plotted in Fig. 3 (squares). As expected, the relationship between f_i and $d_{i,0,S}$ is significant (solid regression line: $d_i = -1.767f_i + 1.541$, $P = 0.026$; Spearman rank correlation, $P = 0.01$). A more surprising result emerged when the decomposition of d_i was used to investigate whether fitness effects measured in yeast are indicative of the evolutionary history of yeast proteins alone, or also of orthologous sequences in the worm. Using the relationship $d_{i,0,C} = d_{i,S,C} - d_{i,0,S}$, we estimated the evolutionary distance from the fungi–animal ancestor to worm. For each of the 48 RBH pairs, the point $(f_i, d_{i,0,C})$ is plotted as a star in Fig. 3 (dashed regression line: $d_i = -1.509f_i + 1.099$, $P = 0.026$; Spearman rank correlation, $P = 6.6 \times 10^{-5}$). Proteins that are less dispensable in yeast have evolved more slowly not only in yeast, but also in the worm. This suggests that non-essential proteins that have measurable effects on yeast fitness are likely to have had proportional effects on worm fitness since the fungi–animal divergence.

How can we reconcile the present finding, namely that relatively dispensable proteins evolve more rapidly, with the previous demonstration that non-essential mouse genes do not evolve more rapidly than essential ones⁵? The 119 yeast proteins plotted in Fig. 1 were deemed non-essential because their deletion did not prevent growth on solid media. In the same study¹³, 356 yeast proteins were found to

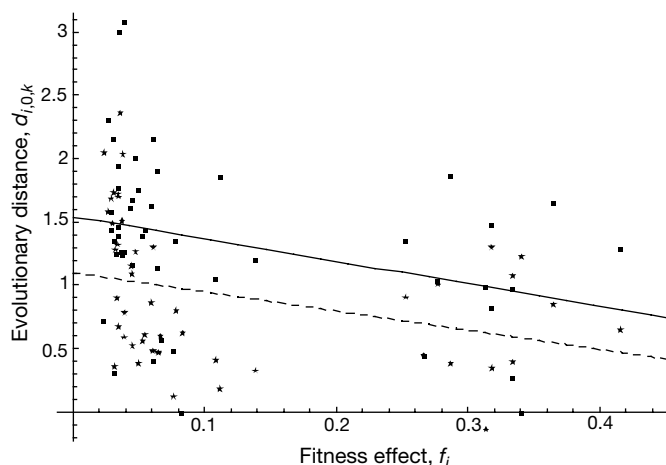


Figure 3 The relationship between fitness effect in yeast and the rate of evolution from the fungi–animal ancestor to yeast or the worm. For 48 proteins, the evolutionary distance from yeast to the worm was decomposed into two segments by the method described in the text. Squares represent distances between yeast and the most recent common ancestor of fungi and animals (solid regression line: $d_i = -1.767f_i + 1.541$, $P = 0.026$; Spearman rank correlation, $P = 0.01$). Stars represent distances between the worm and the most recent common ancestor (dashed regression line: $d_i = -1.509f_i + 1.099$, $P = 0.026$; Spearman rank correlation, $P = 6.6 \times 10^{-5}$). A protein's fitness effect in yeast is predictive of evolutionary rate not only in yeast, but also in the worm.

be essential. Of these, 168 had a RBH in the worm. Using $d_{i,S,C}$ as a measure of evolutionary rate, we compared the evolutionary rates of essential versus non-essential yeast proteins. Consistent with the result from the mouse study, we found no significant difference between the two categories of proteins (Welch's $t = -0.90$, $P = 0.18$; Wilcoxon two-sample, $P = 0.18$). However, when we compared essential proteins with the 'most dispensable' half of non-essential proteins (that is, the 60 proteins with the smallest f_i values), a highly significant difference was observed (Welch's $t = -3.78$, $P = 0.0001$; Wilcoxon two-sample, $P = 7.2 \times 10^{-5}$). Thus, comparison of the evolutionary rates of essential versus non-essential proteins reveals no significant difference, but when we consider a continuum of dispensability, as in Figs 1 and 3, or create by median split a category of the 'least essential' proteins, a significantly higher rate of evolution is observed among dispensable sequences.

This result is readily explained on the basis of the functional form of the relationship between the fitness effect of a protein and the probability of fixation of mutations in its sequence (see Box 1). Assuming that the long-term evolutionary rate of a protein is proportional to the expectation of the probability of fixation of new mutations, which we denote $E(p)$, we are interested in how this expectation varies with fitness effect, that is, in the form of $E(p(f))$. As shown in Fig. 4, the expected probability of fixation falls rapidly as protein fitness effect ranges from 0 to 0.1, but then levels off. This suggests that a protein with a small but measurable fitness effect is,

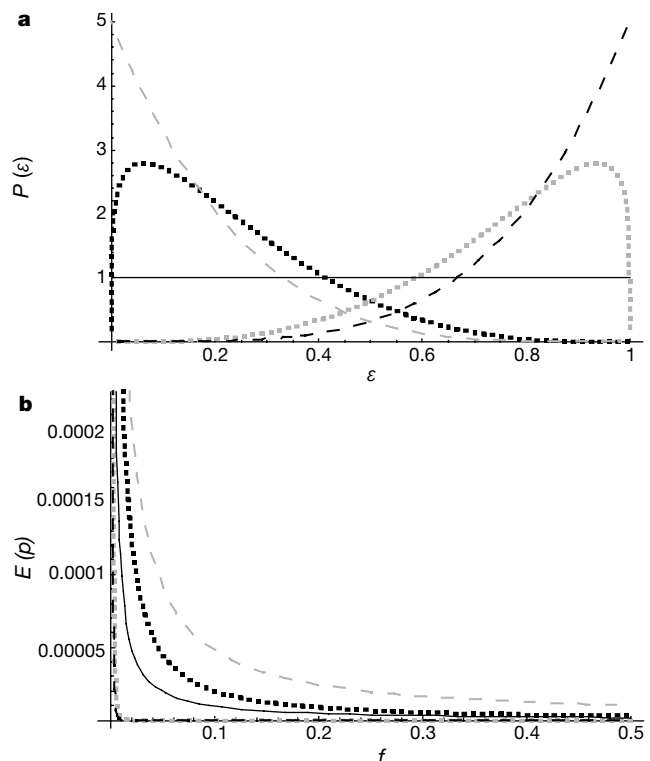


Figure 4 Expected probability of fixation of novel substitutions, $E(p)$, plotted as a function of the fitness effect of a protein, f_i (see Box 1). The L-shape of $E(p(f))$ is robust to changes in our assumptions about $P(\epsilon)$, the probability density function of ϵ . For each form of $P(\epsilon)$ in **a**, $E(p(f))$ is plotted in **b** with a line of corresponding style. The five probability density functions are β distributions with the following parameter values: $(\alpha, \beta) = (1, 2, 4)$, dotted black line; $(4, 1, 2)$, dotted grey line; $(1, 1)$, solid black line; $(5, 1)$, dashed black line; $(1, 5)$, dashed grey line. $E(p(f))$ generally becomes more L-shaped as the product $N_e E(\epsilon)$ increases. To show that the form persists over all reasonable parameter values, we plotted $E(p(f))$ under the assumption of relatively small effective population size ($N_e = 1,000$).

in evolutionary terms, tantamount to essential. As a result, grouping proteins into the categories operative in molecular genetics (essential versus non-essential) obscures the relationship between dispensability and evolutionary rate by including many highly conserved proteins in the non-essential category. In contrast, when we consider a continuum of fitness effects, or a category of proteins with the very smallest effects, a significant relationship between dispensability and evolutionary rate is revealed. This explanation is also consistent with recent suggestions that genes in which mutations produce no observable phenotypic changes (a classification analogous to our category of proteins of smallest effect) are less likely to have homologues in other taxa than are genes in which mutations have any detectable effects^{19,20}.

In addition to confirming a fundamental prediction about protein evolution^{1–3}, these results suggest that genome-wide genetic

studies, such as parallel analysis¹³ and footprinting²¹, are relevant to evolutionary study, despite the artificiality of the laboratory environment. High-throughput measurement of fitness effect is sufficiently indicative of a protein's dispensability over the last billion years for the relationships we observe in Figs 1 and 3 to be highly significant. The significance and phylogenetic conservation of the relationship between dispensability and evolutionary rate is particularly surprising in view of the number of processes that would tend to obscure the relationship over long stretches of evolutionary time. Changes in the relative importance of individual proteins to the organism, the evolution of functional robustness against mutations, and compensatory mutations masking deleterious ones would all tend, over evolutionary time, to obscure the relationship that we observe. Among the proteins we considered, these processes have evidently been sufficiently slow to allow detection of the relationship between protein dispensability and evolutionary rate. □

Box 1

Fitness effect of a protein and probability of fixation of a sequence mutation

Consider a standard Wright–Fisher model of drift combined with viability selection against deleterious mutations^{27,28}. We define the model's fitness scheme in a fashion that is compatible with the experimental measurement of single-gene deletion effects in yeast: individuals homozygous for a new mutation have fitness $1 - \varepsilon f$, where f is the protein's fitness effect, as defined above, and ε is the fitness effect of a specific amino acid substitution at a single site, expressed as a fraction of the effect of deleting the gene altogether. Thus, for those substitutions that render a protein dysfunctional, $\varepsilon = 1$; for those that have no effect on protein function, $\varepsilon = 0$; and for those that result in partial loss of function, $0 < \varepsilon < 1$. We assume that the distribution of ε is described by the probability density function $P(\varepsilon)$, which is independent of f . Individuals homozygous for a wild-type allele have fitness 1, and heterozygotes have fitness $1 - h\varepsilon f$, where h represents the degree of dominance. By substituting this fitness scheme into the Wright–Fisher model and following Kimura's derivation of fixation probability^{27,28}, we observe that in a population of effective size N_e the probability of fixation of a novel substitution is given by

$$p = \frac{\int_0^{1/2N} \psi(x) dx}{\int_0^1 \psi(x) dx} \quad (1)$$

where $\psi(x) = \exp[2N_e f \varepsilon x (2h + (1 - 2h)x)]$. Under the assumption that the long-term evolutionary rate of a protein is roughly proportional to $E(p)$, the expectation of the probability of fixation of a new substitution, we examined the functional form of $E(p(f))$. Numerical analysis of equation (1) allowed us to eliminate the parameter h from detailed consideration. Although the degree of dominance can have a significant impact on the probability of fixation of a given mutation, its effect on the functional form of $E(p(f))$ is small. To identify regions of parameter space in which h has the greatest effect on the form of $E(p(f))$, we maximized $\int_0^1 (p(N_e, \varepsilon, f, h_1) - p(N_e, \varepsilon, f, h_2)) (p(N_e, \varepsilon, f, h_1) + p(N_e, \varepsilon, f, h_2))^2 df$ in the parameter space bounded by $1 \leq N_e \leq 10^5$, with respect to pairs of h_1 and h_2 on $[0, 0.5]$. Even in the regions identified, the impact of h on the form of $E(p(f))$ remained negligible. For the following analysis, we adopted the value of $h = 0.1$. What form the probability density function of ε , $P(\varepsilon)$, should take is not known. It has been suggested that the distribution of evolutionary rate across sites is described by gamma density with shape parameter between 0.3 and 0.8 (refs 23 and 29). If we assume that sites vary in their evolutionary rate because they have different expected values of ε (that is, certain sites are consistently associated with higher values of ε), and that $\int_0^1 P(\varepsilon) d\varepsilon$ is indicative of the fraction of sites in which $a \leq \varepsilon \leq b$, then we can expect $P(\varepsilon)$ to resemble $\beta(5, 1)$ (Fig. 4a dashed black line): this is the probability density function of ε that results in a distribution of p that resembles gamma density. However, because the evidence in support of $\beta(5, 1)$ is based on non-trivial assumptions, we consider several different probability density functions, showing that the L-shaped functional form of $E(p(f))$ is robust (Fig. 4).

Methods

If a query of all worm open reading frames (ORFs) with yeast ORF i yielded the set of hits $\{W\}$ with $P < 10^{-10}$ and more than 80% sequence alignment, and a query of all yeast ORFs with worm ORF j yielded the set of hits $\{Y\}$ with $P < 10^{-10}$ and more than 80% sequence alignment, then the sequence pair (ORF i , ORF j) was included in the analysis only if i was the member of $\{Y\}$ with the lowest P value and j was the member of $\{W\}$ with the lowest P value.

The number of substitutions per site, d , was estimated by two different methods. In method 1 (ref. 22), we numerically solved $q = \ln(1 + 2d)/2d$, where q is the proportion of identical sites between aligned sequences. In method 2 (ref. 23), we numerically solved $[(q - q_e)/(1 - q_e)] = \int_0^\infty (\alpha^\alpha / \Gamma(\alpha)) x^{\alpha-1} e^{-(\alpha+d)x} dx$, where q_e is the expected identity for random sequences, and $\alpha = 0.31$ is the parameter of the gamma density function describing evolutionary rate variation among sites. For decomposition of $d_{i,S,C}$ into $d_{i,0,S}$ and $d_{i,0,C}$, we relied on method 1 for two reasons. First, it has been used and tested by others^{24,25}, whereas method 2 has had less verification. Second, a number of results in the course of this work suggested that method 2 may overestimate large distances. Outgroup sequences were identified in the Clusters of Orthologous Groups database²⁶ and were distributed among the following species: *A. fulgidus*, *M. jannaschii*, *M. thermoautotrophicum*, *P. horikoshii*, *A. aeolicus*, *T. maritima*, *Synechocystis* spp., *E. coli*, *B. subtilis*, *M. tuberculosis*, *H. influenzae*, *H. pylori*, *M. genitalium*, *B. burgdorferi*, *T. pallidum*, *Chlamydia* spp. and *R. prowazekii*. An estimate of $d_{i,0,S}$ was determined using each outgroup for which an orthologue of i was available, and these estimates were averaged.

Received 4 January; accepted 10 May 2001.

- Ohta, T. The nearly neutral theory of molecular evolution. *Annu. Rev. Ecol. Syst.* **23**, 263–286 (1992).
- Ohta, T. Slightly deleterious mutant substitutions in evolution. *Nature* **246**, 96–98 (1973).
- Wilson, A. C., Carlson, S. S. & White, T. J. Biochemical evolution. *Annu. Rev. Biochem.* **46**, 573–639 (1977).
- Wagner, A. Robustness against mutations in genetic networks of yeast. *Nature Genet.* **24**, 355–361 (2000).
- Hurst, L. D. & Smith, N. G. C. Do essential genes evolve slowly? *Curr. Biol.* **9**, 747–750 (1999).
- Kuma, K., Iwabe, N. & Miyata, T. Functional constraints against variations on molecules from the tissue level—slowly evolving brain-specific genes demonstrated by protein-kinase and immunoglobulin supergene families. *Mol. Biol. Evol.* **12**, 123–130 (1995).
- Williams, E. J. B. & Hurst, L. D. The proteins of linked genes evolve at similar rates. *Nature* **407**, 900–903 (2000).
- Hughes, A. L. Rapid evolution of immunoglobulin superfamily C2 domains expressed in immune system cells. *Mol. Biol. Evol.* **14**, 1–5 (1997).
- Stockbauer, K. E. et al. Hypervariability generated by natural selection in an extracellular complement-inhibiting protein of serotype M1 strains of group A *Streptococcus*. *Proc. Natl Acad. Sci. USA* **95**, 3128–3133 (1998).
- Yamaguchi, Y. & Gojobori, T. Evolutionary mechanisms and population dynamics of the third variable envelope region of HIV within single hosts. *Proc. Natl Acad. Sci. USA* **94**, 1264–1269 (1997).
- Makalowski, W. & Boguski, M. S. Synonymous and nonsynonymous substitution distances are correlated in mouse and rat genes. *J. Mol. Evol.* **47**, 119–121 (1998).
- Tourasse, N. J. & Li, W. H. Selective constraints, amino acid composition, and the rate of protein evolution. *Mol. Biol. Evol.* **17**, 656–664 (2000).
- Wenzler, E. A., Shoemaker, D. D., Astromoff, A. & Liang, H. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901–906 (1999).
- Rivera, M. C., Jain, R., Moore, J. E. & Lake, J. A. Genomic evidence for two functionally distinct gene classes. *Proc. Natl Acad. Sci. USA* **95**, 6239–6244 (1998).
- Chervitz, S. A. et al. Comparison of the complete protein sets of worm and yeast: orthology and divergence. *Science* **282**, 2022–2028 (1998).
- Robinson, M., Gouy, M., Gautier, C. & Mouchiroud, D. Sensitivity of the relative-rate test to taxonomic sampling. *Mol. Biol. Evol.* **15**, 1091–1098 (1998).
- Sarich, V. M. & Wilson, A. C. Generation time and genomic evolution in primates. *Science* **179**, 1144–1147 (1973).
- Wu, C.-I. & Li, W.-H. Evidence for higher rates of nucleotide substitution in rodents than in man. *Proc. Natl Acad. Sci. USA* **82**, 1741–1745 (1985).
- Ashburner, M. et al. An exploration of the sequence of a 2.9-Mb region of the genome of *Drosophila melanogaster*: The *Adh* region. *Genetics* **153**, 179–219 (1999).

20. Fraser, A. G. *et al.* Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature* **408**, 325–330 (2000).
21. Smith, V., Chou, K. N., Lashkari, D., Botstein, D. & Brown, P. O. Functional analysis of the genes of yeast chromosome V by genetic footprinting. *Science* **274**, 2069–2074 (1996).
22. Grishin, N. V. Estimation of the number of amino acid substitutions per site when the substitution rate varies among sites. *J. Mol. Evol.* **41**, 675–679 (1995).
23. Grishin, N. V., Wolf, Y. I. & Koonin, E. V. From complete genomes to measures of substitution rate variability within and between proteins. *Genome Res.* **10**, 991–1000 (2000).
24. Feng, D. & Doolittle, R. Converting amino acid alignment scores into measures of evolutionary time: A simulation study of various relationships. *J. Mol. Evol.* **44**, 361–370 (1997).
25. Huynen, M. & Bork, P. Measuring genome evolution. *Proc. Natl Acad. Sci. USA* **95**, 5849–5856 (1998).
26. Tatusov, R. L., Galperin, M. Y., Nalale, D. A. & Koonin, E. V. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **28**, 33–36 (2000).
27. Ewens, W. H. *Mathematical Population Genetics* (eds Krickeberg, K. & Levin, S. A.) (Springer, New York, 1979).
28. Kimura, M. On the probability of fixation of mutant genes in a population. *Genetics* **47**, 713–719 (1962).
29. Zhang, J. & Gu, X. Correlation between the substitution rate and rate variation among sites in protein evolution. *Genetics* **149**, 1615–1625 (1998).

Acknowledgements

We thank D. Petrov and M. Feldman for much guidance and support. M. Cherry and A. Chu provided assistance with genomic and functional genomic data.

Correspondence and requests for materials should be addressed to A.E.H. (e-mail: aehirsh@stanford.edu).

The contribution of sensory experience to the maturation of orientation selectivity in ferret visual cortex

Leonard E. White^{*†}, David M. Coppola[‡] & David Fitzpatrick^{*}

^{*} Department of Neurobiology, [†] Department of Community and Family Medicine, Division of Physical Therapy, Duke University Medical Center, Durham, North Carolina 27710, USA

[‡] Department of Biology, Centenary College, Shreveport, Louisiana 71134, USA

Sensory experience begins when neural circuits in the cerebral cortex are still immature; however, the contribution of experience to cortical maturation remains unclear. In the visual cortex, the selectivity of neurons for oriented stimuli at the time of eye opening is poor^{1–5} and increases dramatically after the onset of visual experience^{3–8}. Here we investigate whether visual experience has a significant role in the maturation of orientation selectivity and underlying cortical circuits^{9–12} using two forms of deprivation: dark rearing, which completely eliminates experience, and binocular lid suture, which alters the pattern of sensory driven activity¹³. Orientation maps were present in dark-reared ferrets, but fully mature levels of tuning were never attained. In contrast, only rudimentary levels of orientation selectivity were observed in lid-sutured ferrets. Despite these differences, horizontal connections in both groups were less extensive and less clustered than normal, suggesting that long-range cortical processing is not essential for the expression of orientation selectivity, but may be needed for the full maturation of tuning. Thus, experience is beneficial or highly detrimental to cortical maturation, depending on the pattern of sensory driven activity.

The interactions between visually driven activity and endogenous mechanisms of development that shape the maturation of orientation selectivity remain unclear. Some studies suggest that visual experience is not required for the full maturation of selectivity^{6,7,14,15}, whereas others propose a more constructive influence^{3,8,16}. To clarify these interactions, we first asked whether the rapid maturation of

orientation selectivity that follows eye opening could be attributed solely to endogenous developmental mechanisms. We analysed the development of orientation selectivity in 16 ferrets that were dark reared from about two weeks before eye opening and brought into the light only after the induction of anaesthesia for the terminal experiment. Representative results are shown in Fig. 1. About one week after eye opening, this dark-reared ferret showed considerable maturation of orientation selectivity, despite the complete absence of previous visual experience (see Fig. 1b). Qualitatively, the map of orientation preference was very similar to normal in its areal extent and internal structure: orientation preference was organized in a continuous fashion except for point discontinuities around which preference was mapped in pinwheel-like arrangements (see insets in Fig. 1). However, careful examination of the difference images indicated that the responses were less selective than normal, as judged by the reduced contrast in the optical maps compared with normal (compare Fig. 1a with b).

This impression was substantiated by a quantitative analysis of cardinal difference images (that is, horizontal versus vertical) in a large series of normal and deprived ferrets that sampled the period of orientation map development. In normal animals, differential optical responses to orthogonal angles were first seen at or just before the time of natural eye opening (postnatal days 30–35), as reported previously^{4,5} (Fig. 2). The selectivity of these responses increased greatly during the next two weeks and achieved a plateau by the eighth or ninth week (Fig. 2b). Ferrets that were raised in complete darkness showed evidence of a similar progression from before the time of eye opening through to the sixth postnatal week; thereafter, the degree of selectivity was significantly less than normal. Although cortical responses can be evoked through naturally closed eyelids in very young ferrets (postnatal weeks 4–5)¹⁷, our results indicate that the contribution of visual experience to the maturation of orientation selectivity is confined to the developmental period that follows eye opening.

These data show that experience-independent mechanisms have the capacity to establish the map of orientation preference and promote its maturation. However, they also demonstrate that this capacity is limited: fully mature levels of selectivity were achieved only with the benefit of normal experience. This may suggest that orientation-selective responses are determined primarily by endogenous factors and that the influence of visually driven activity is limited to sharpening orientation tuning. On the other hand, it is possible that visually driven activity has a more important role, one that cannot be inferred from experiments in which its influence has been completely eliminated. To address this possibility, we analysed the development of orientation selectivity in ferrets that were subjected to binocular lid suture, which does not prevent light activation of the retina but does alter experience by passing only very low spatial and temporal frequencies¹³.

In contrast to the more modest effects of dark rearing, lid suture had devastating effects on the maturation of orientation selectivity. In the representative case shown in Fig. 1c, there was very little evidence of differential cortical response to orthogonal orientations. Notably, the cortex had not simply become unresponsive; single-condition optical imaging (that is, stimulus versus blank) showed robust but much less modular cortical activation to each stimulus. The group analysis indicated that unlike the dark-reared animals, lid-sutured ferrets were similar to normal only at the time of natural eye opening (see Fig. 2 and Supplementary Information). Thereafter, lid-sutured ferrets failed to show any improvement in selectivity; indeed, this group of ferrets showed a steady decline with ongoing deprivation. To be sure that the effects of lid suture were related to altered patterns of neural activity rather than some non-specific consequence of lid suture, one ferret was lid-sutured and reared in total darkness. The map of orientation preference developed in this animal to the same degree as an unsutured littermate that was imaged on the same day (see Supplementary Information).

The maturation of the map of orientation preference proceeded to a limited extent in the absence of vision, but not with abnormal experience engendered by vision through sutured eyelids. This effect of lid suture differs from reports in kittens in which binocular lid suture did not prevent the maturation of the map, only the long-term maintenance of adult levels of selectivity^{6,7}. These discrepancies may be explained by species differences in the timing of susceptibility to the influence of experience or, more simply, by differences in the amount of light transmitted through sutured eyelids.

We next sought to confirm our results with electrophysiological

recordings in a subset of the same animals that were used for imaging. The electrophysiological assessments of orientation selectivity were consistent with the results obtained by optical imaging: dark-reared ferrets showed an intermediate level of tuning that was sharper than the lid-sutured group but not as selective as that seen in normal ferrets (see Supplementary Information). The mean values of an orientation selectivity index for electrophysiological data (OSI_{EP}) ($\pm$ standard error) for each of three groups were: normal = 66.2 ± 2.7 ; dark-reared = 44.0 ± 3.8 ; and lid-sutured = 36.4 ± 4.3 ($F_{2,76} = 21.3$, $P < 0.001$).

Finally, we asked whether the rapid maturation of orientation

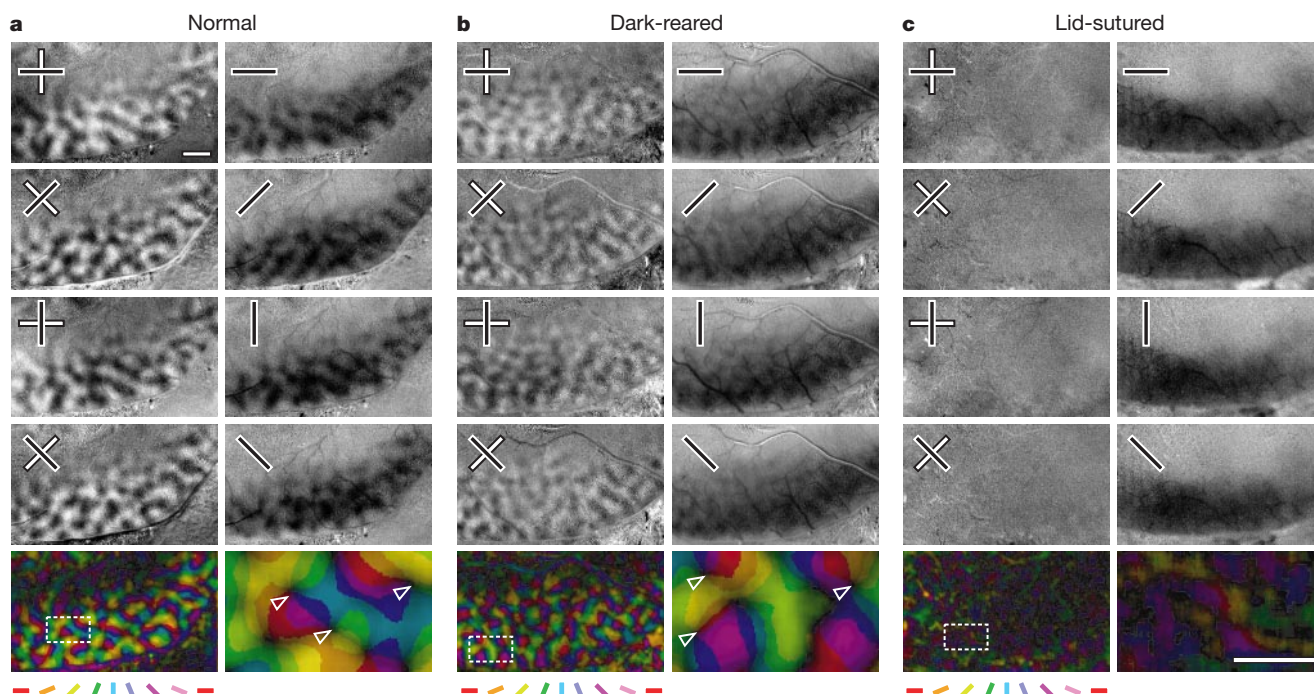


Figure 1 Maps of orientation preference in normal and visually deprived ferrets.

a, Difference images (left column) show selective responses to orthogonal gratings in a 39-day-old ferret reared normally. The same patterns are seen in the adjacent single-condition images. The overall organization is shown in the polar-magnitude map where

colour represents preference and brightness represents selectivity. Pinwheel centres (white arrowheads) are shown at higher magnification to the right. **b**, Images from a 37-day-old, dark-reared ferret. **c**, Images from a 43-day-old, lid-sutured ferret. Note the absence of selectivity despite the strong visual response. Scale bars: 1 mm.

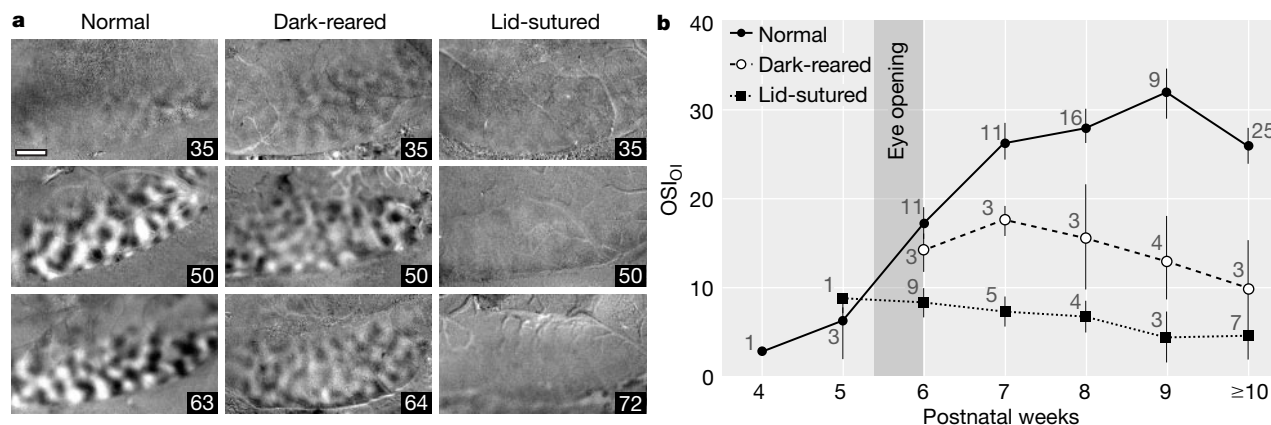


Figure 2 Quantitative assessment of orientation preference in normal and visually deprived ferrets. **a**, Cardinal difference images show the increase in selectivity seen in normal ferrets after eye opening, the substantial development of the map in dark-reared ferrets, and the devastating effects of lid suture (numbers indicate postnatal age; scale bar: 1 mm). **b**, Plot of OSI_{OI} (mean $\pm$ standard error) versus postnatal age for each

treatment group (numbers next to symbols indicate sample size). After week six, rearing condition was significant ($F_{2,99} = 3.3$, $P < 0.05$), as was the condition/age interaction ($F_{8,99} = 3.6$, $P < 0.001$) with significant differences among all conditions ($P < 0.01$, *post hoc* Tukey–Kramer Honestly Significant Difference).

selectivity that follows eye opening and its disruption by visual deprivation could be understood in terms of the development of the horizontal network in layer 2/3. This system of long-range, clustered connections that link together cortical columns that share similar orientation preferences^{9–12} has been proposed to be involved in the establishment and maturation of the map of orientation preference^{10,18}. Moreover, previous studies in cats demonstrated that binocular lid suture prevents the expression of the mature pattern of horizontal connectivity^{19,20}. Our results confirm this conclusion: axonal distributions in lid-sutured ferrets gave rise to fewer terminal clusters and covered less cortical area, in comparison with normal (Fig. 3). However, our data show that the devastating effects of lid suture on orientation selectivity cannot be attributed solely to alterations in the horizontal network, as these connections are also severely affected in dark-reared animals. This effect is different from the results of an earlier study in which horizontal connections appeared normal in two dark-reared kittens²¹. This inconsistency may be due to methodological differences (tracer substance, injection size, sample size) and/or differences in the timing of horizontal connection outgrowth relative to the onset of visual experience. Our results suggest that the restricted distribution of horizontal connections underlies the effects of dark rearing on orientation preference, whereas the more devastating effects of lid suture are explained by alterations in the functional organization of other cortical circuits.

The role of experience in the development of orientation selec-

tivity has been the subject of numerous studies with results and interpretations that often conflict^{13,22,23}. However, most recent accounts distinguish two phases of maturation: an initial period of experience-independent development in which the map of orientation preference emerges and selectivity reaches maturity, followed by a second experience-dependent phase in which experience is required to consolidate and maintain orientation-selective responses^{6,7,14,15,24}. Our results showing that experience exerts a powerful influence during the maturational phase of orientation development are difficult to reconcile with this view. In the ferret, adult levels of orientation tuning are achieved only with the benefit of visual experience. Our results emphasize that the pattern of visually driven activity, not merely its presence, is a critical factor in promoting the complete maturation of orientation selectivity. Indeed, abnormal patterns of visually driven activity are far more disruptive to the maturation of orientation-selective responses than the total absence of visually driven activity. Additional evidence to support the disruptive influence of altered patterns of retinal activity has come from studies using electrical or pharmacological manipulations^{25,26}. Although it is unclear whether these interventions altered developmental events during the initial establishment of orientation selectivity or during the later phase of maturation (or both), our results demonstrate that disruptions in the pattern of visually driven activity during the maturational phase alone would be sufficient to explain these observations.

But how might neural activity engendered by normal and

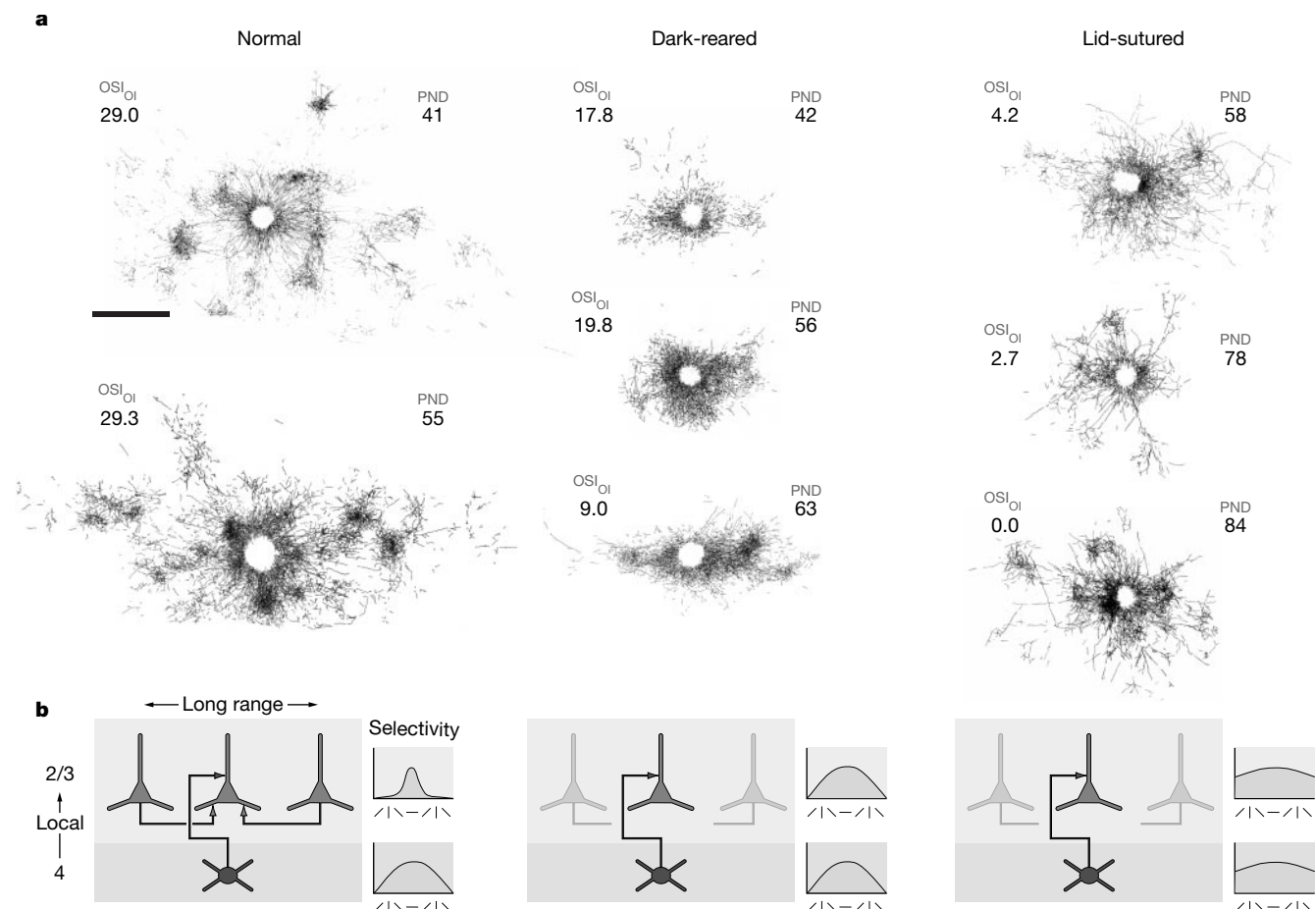


Figure 3 Horizontal connections in layer 2/3 of normal and visually deprived ferrets. **a**, Representative cases: the vacant area in the centre of each tracing contains the injection site (scale bar: 1 mm). PND, postnatal day. **b**, Models for the effects of experience on columnar and long-range processing. Normal experience sharpens orientation tuning within layer 2/3 (see ref. 4). Dark rearing prevents the full elaboration of

horizontal connections and the increased selectivity that requires long-range computations. Horizontal connections are similarly impaired with lid suture and altered patterns of visually driven activity interfere with the establishment of orientation selectivity in layer 4.

abnormal visual experience interact with endogenous factors to produce sharpening or broadening, respectively, of orientation selectivity? Current models propose that intrinsic patterns of activity, in particular the local correlation structure of converging on-centre and off-centre inputs from the lateral geniculate nucleus, are sufficient to establish the orientation-selective responses of neurons in cortical layer 4 (reviewed in ref. 14). We suggest that the endogenous activities of on-centre and off-centre inputs²⁷ shape local columnar circuitry and account for the establishment and partial maturation of orientation selectivity in the total absence of visual experience, without the benefit of long-range, clustered horizontal connections (see Fig. 3b). However, lighted rearing conditions allow for interactions between endogenous factors and sensory driven activity. Given the prevalence of long contours in the visual environment, normal experience would be expected to act synergistically with intrinsic activity and other endogenous factors to increase the temporal and spatial precision of orientation-biased correlations, thereby sharpening cortical orientation circuits by a hebbian process. This should occur both within local columnar circuits and across the horizontal network in layer 2/3; indeed, our anatomical data show that the full elaboration of this network requires the influence of normal visual experience. In contrast, abnormal experience through light-scattering eyelids should abrogate this synergy by co-activating inputs over a broader spatial scale, thus increasing the correlation of inputs with different centre types that would otherwise tend to segregate. Consequently, both intra- and inter-columnar stages of orientation processing would be impaired, and the rudimentary map of orientation preference present at the time of eye opening fails to mature and progressively weakens. In conclusion, the pattern of activity evoked by visual experience affects the architecture of intracortical circuits and is a critical factor in the maturation of orientation selectivity in the developing visual cortex. □

Methods

All experimental procedures were approved by the Duke University Institutional Animal Care and Use Committee and were performed in compliance with guidelines published by the National Institutes of Health (USA).

Rearing conditions

Normal sable ferrets (*Mustela putorius furo*) of both sexes ($n = 76$) were reared in a 12 h light/dark cycle. Other ferrets ($n = 16$) were subjected to continual, absolute dark rearing beginning on postnatal day 21 or 23 (with day of birth being postnatal day 0) to maximize exposure of the mother and kits to normal circadian cycles, while ensuring the onset of deprivation before the development of reliable responses in cortical neurons⁴¹⁷. Between postnatal days 21–27, 29 ferrets were briefly anaesthetized and subjected to binocular lid suture.

Optical imaging

Optical imaging of intrinsic signals was performed with an enhanced video acquisition system (Optical Imaging), as described previously^{12,28}. The stimuli were moving, high-contrast rectangular wave gratings (0.1 cycle per degree) oriented at 0, 45, 90 or 135°. Difference images were generated by subtracting the optical responses to presentation of each member of a pair of orthogonal gratings. To compare different animals, an index of orientation selectivity for optical images (OSI_{OI}) was computed from the cardinal difference images by first clipping at ± 3 s.d. from the median and then calculating the s.d. of the distribution of greylevels within V1 and V2. Mean OSI_{OI} values binned by postnatal week were analysed using two-way analysis of variance, with rearing condition and age as the main effects (JMP Statistical Software). To assess cortical responsiveness, single condition images were generated by subtracting the response to a blank screen from the response to a single grating.

Electrophysiology

In a subset of cases (38 sites from 28 normal ferrets, 25 sites from 6 dark-reared ferrets, and 16 sites from 5 lid-sutured ferrets), a tungsten microelectrode (impedance = 8–14 M Ω) was inserted into selected cortical sites. Multi-unit activity was recorded from cortical layer 3 and the signals discriminated and tabulated using Spike2 software (Cambridge Electronic Design). For each recording site, the optimum orientation preference of the multi-unit activity was determined by panning an orientated bar ($40^\circ \times 0.4^\circ$) across the receptive field of the site in one of 9 or 18 orientations. The average spike counts from five replications were used to construct histograms, which were then fit by a gaussian distribution using a modified downhill simplex method²⁹. We then calculated an

orientation selectivity index for electrophysiological data (OSI_{EP}):

$$OSI_{EP} = [P2/(P1 + P2)] \times [1 - (P4/\text{number of stimuli})] \times 100$$

where $P1$ is baseline, $P2$ is peak height minus baseline, and $P4$ is 0.83 times half-width at half-height.

Anatomy

After optical imaging, 21 normal ferrets, five dark-reared ferrets and five hemispheres from three lid-sutured ferrets each received a small iontophoretic injection of biocytin into the visual cortex, as previously described¹². Labelled axons in every tangential section through layers 2/3 were traced, and the areal extent of the composite distribution and the number of terminal clusters were measured using NIH Image software (developed at the US National Institutes of Health; available at <http://rsb.info.nih.gov/nih-image/>).

Received 17 January; accepted 4 May 2001.

- Wiesel, T. N. & Hubel, D. H. Ordered arrangement of orientation columns in monkeys lacking visual experience. *J. Comp. Neurol.* **158**, 307–318 (1974).
- Hubel, D. H. & Wiesel, T. N. Receptive fields of cells in striate cortex of very young, visually inexperienced kittens. *J. Neurophysiol.* **26**, 994–1002 (1963).
- Frégnac, Y. & Imbert, M. Early development of visual cortical cells in normal and dark-reared kittens: the relationship between orientation selectivity and ocular dominance. *J. Physiol.* **278**, 27–44 (1978).
- Chapman, B. & Stryker, M. P. Development of orientation selectivity in ferret visual cortex and effects of deprivation. *J. Neurosci.* **13**, 5251–5262 (1993).
- Chapman, B., Stryker, M. P. & Bonhoeffer, T. Development of orientation preference maps in ferret primary visual cortex. *J. Neurosci.* **16**, 6443–6453 (1996).
- Gödecke, I., Kim, D. S., Bonhoeffer, T. & Singer, W. Development of orientation preference maps in area 18 of kitten visual cortex. *Eur. J. Neurosci.* **9**, 1754–1762 (1997).
- Crair, M. C., Gillespie, D. C. & Stryker, M. P. The role of visual experience in the development of columns in cat visual cortex. *Science* **279**, 566–570 (1998).
- Blakemore, C. & Van Sluyters, R. C. Innate and environmental factors in the development of the kitten's visual cortex. *J. Physiol.* **248**, 663–716 (1975).
- Gilbert, C. D. & Wiesel, T. N. Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex. *J. Neurosci.* **9**, 2432–2442 (1989).
- Durack, J. C. & Katz, L. C. Development of horizontal projections in layer 2/3 of ferret visual cortex. *Cereb. Cortex* **6**, 178–183 (1996).
- Ruthazer, E. S. & Stryker, M. P. The role of activity in the development of long-range horizontal connections in area 17 of the ferret. *J. Neurosci.* **16**, 7253–7269 (1996).
- Bosking, W. H., Zhang, Y., Schofield, B. & Fitzpatrick, D. Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *J. Neurosci.* **17**, 2112–2127 (1997).
- Sherman, S. M. & Spear, P. D. Organization of visual pathways in normal and visually deprived cats. *Physiol. Rev.* **62**, 738–855 (1982).
- Chapman, B., Gödecke, I. & Bonhoeffer, T. Development of orientation preference in the mammalian visual cortex. *J. Neurobiol.* **41**, 18–24 (1999).
- Miller, K. D., Erwin, E. & Kayser, A. Is the development of orientation selectivity instructed by activity? *J. Neurobiol.* **41**, 44–57 (1999).
- Sengpiel, F., Stawinski, P. & Bonhoeffer, T. Influence of experience on orientation maps in cat visual cortex. *Nature Neurosci.* **2**, 727–732 (1999).
- Krug, K., Akerman, C. J. & Thompson, I. D. Responses of neurons in neonatal cortex and thalamus to patterned visual stimulation through the naturally closed lids. *J. Neurophys.* **85**, 1436–1443 (2001).
- Shouval, H. Z., Goldberg, D. H., Jones, J. P., Beckerman, M. & Cooper, L. N. Structured long-range connections can provide a scaffold for orientation maps. *J. Neurosci.* **20**, 1119–1128 (2000).
- Callaway, E. M. & Katz, L. C. Effects of binocular deprivation on the development of clustered horizontal connections in cat striate cortex. *Proc. Natl Acad. Sci. USA* **88**, 745–749 (1991).
- Zufferey, P. D., Jin, F., Nakamura, H., Tettoni, L. & Innocenti, G. M. The role of pattern vision in the development of cortico-cortical connections. *Eur. J. Neurosci.* **11**, 2669–2688 (1999).
- Lübke, J. & Albus, K. Rapid rearrangement of intrinsic tangential connections in the striate cortex of normal and dark-reared kittens: lack of exuberance beyond the second postnatal week. *J. Comp. Neurol.* **323**, 42–58 (1992).
- Frégnac, Y. & Imbert, M. Development of neuronal selectivity in primary visual cortex of cat. *Physiol. Rev.* **64**, 325–434 (1984).
- Daw, N. *Visual Development* (Plenum, New York, 1995).
- Mower, G. D., Berry, D., Burchfiel, J. L. & Duffy, F. H. Comparison of the effects of dark-rearing and binocular suture on development and plasticity of cat visual cortex. *Brain Res.* **220**, 255–267 (1981).
- Weliky, M. & Katz, L. C. Disruption of orientation tuning in visual cortex by artificially correlated neuronal activity. *Nature* **386**, 680–685 (1997).
- Chapman, B. & Gödecke, I. Cortical cell orientation selectivity fails to develop in the absence of ON-center retinal ganglion cell activity. *J. Neurosci.* **20**, 1922–1930 (2000).
- Wong, R. O. & Oakley, D. M. Changing patterns of spontaneous bursting activity of on and off retinal ganglion cells during development. *Neuron* **16**, 1087–1095 (1996).
- Bonhoeffer, T. & Grünwald, A. in *Brain Mapping: The Methods* (ed. Toga, A. W. & Mazziotta, J. C.) 55–97 (Academic, San Diego, 1996).
- Bosking, W. H., Kretz, R., Pucak, M. L. & Fitzpatrick, D. Functional specificity of callosal connections in tree shrew striate cortex. *J. Neurosci.* **20**, 2346–2359 (2000).

Supplementary information is available at Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy form the London editorial office of Nature.

Acknowledgements

We thank B. Bosking and J. Crowley for discussions.

Correspondence and requests for materials should be addressed to L.E.W. (e-mail: white033@mc.duke.edu).

A defined range of guard cell calcium oscillation parameters encodes stomatal movements

Gethyn J. Allen*, Sarah P. Chu*, Carrie L. Harrington*, Karin Schumacher†, Thomas Hoffmann‡, Yat Y. Tang*, Erwin Grill‡ & Julian I. Schroeder*

* Cell and Developmental Biology Section, Division of Biology and Center for Molecular Genetics, University of California, San Diego, La Jolla, California 92093-0116, USA

† ZMBP-Pflanzenphysiologie, Universität Tübingen, Auf der Morgenstelle 1, 72076 Tübingen, Germany

‡ Technische Universität, München Lehrstuhl für Botanik, D-85350 Freising-Weihenstephan, Germany

Oscillations in cytosolic calcium concentration ($[Ca^{2+}]_{cyt}$) are central regulators of signal transduction cascades¹, although the roles of individual $[Ca^{2+}]_{cyt}$ oscillation parameters in regulating downstream physiological responses remain largely unknown. In plants, guard cells integrate environmental and endogenous signals to regulate the aperture of stomatal pores² and $[Ca^{2+}]_{cyt}$ oscillations are a fundamental component of stomatal closure^{3,4}. Here we systematically vary $[Ca^{2+}]_{cyt}$ oscillation parameters in *Arabidopsis* guard cells using a 'calcium clamp'^{3,5-7} and show that $[Ca^{2+}]_{cyt}$ controls stomatal closure by two mechanisms. Short-term 'calcium-reactive' closure occurred rapidly when $[Ca^{2+}]_{cyt}$ was elevated, whereas the degree of long-term steady-state closure was 'calcium programmed' by $[Ca^{2+}]_{cyt}$ oscillations within a defined range of frequency, transient number, duration and

amplitude. Furthermore, in guard cells of the *gca2* mutant⁸, $[Ca^{2+}]_{cyt}$ oscillations induced by abscisic acid and extracellular calcium had increased frequencies and reduced transient duration, and steady-state stomatal closure was abolished. Experimentally imposing $[Ca^{2+}]_{cyt}$ oscillations with parameters that elicited closure in the wild type restored long-term closure in *gca2* stomata. These data show that a defined window of guard cell $[Ca^{2+}]_{cyt}$ oscillation parameters programs changes in steady-state stomatal aperture.

In animal cells, $[Ca^{2+}]_{cyt}$ oscillations can regulate the activation of downstream components of signalling cascades^{5-7,9-12}. In plant cells, $[Ca^{2+}]_{cyt}$ oscillations occur in response to many stimuli^{3,13-17}, but hypotheses that oscillation kinetics control physiological responses in plants^{4,18} have not been systematically investigated. We incubated *Arabidopsis* guard cells expressing the green-fluorescent-protein-based calcium indicator yellow cameleon 2.1 (see Methods)^{19,20} for 2.5 h in the light in a high KCl (100 mM) buffer to elicit stomatal opening. These guard cells showed a stable, low resting $[Ca^{2+}]_{cyt}$ level, and stomatal aperture remained constant for the subsequent 3 h (see Supplementary Information Fig. 1).

The influx of extracellular calcium ($[Ca^{2+}]_{ext}$) into guard cells is enhanced by plasma membrane hyperpolarization^{3,21-24}. $[Ca^{2+}]_{cyt}$ oscillations were imposed in guard cells by repeatedly changing between the high (100 mM) KCl buffer (depolarizing buffer) and a low (0.1 mM) KCl buffer containing 10 mM calcium (hyperpolarizing buffer)³ (Fig. 1). Exchanges between the depolarizing and hyperpolarizing buffers every 5 min created oscillations in guard cell $[Ca^{2+}]_{cyt}$ with a fixed 10-min hyperpolarization to hyperpolarization period (Fig. 1a). Continuous repetitive exchanges induced 5–9 transients, depending on the degree of attenuation in individual guard cells (six shown in Fig. 1a, bottom panel). Following the imposed oscillations, guard cells were returned to the depolarizing buffer under white light, and the steady-state apertures were

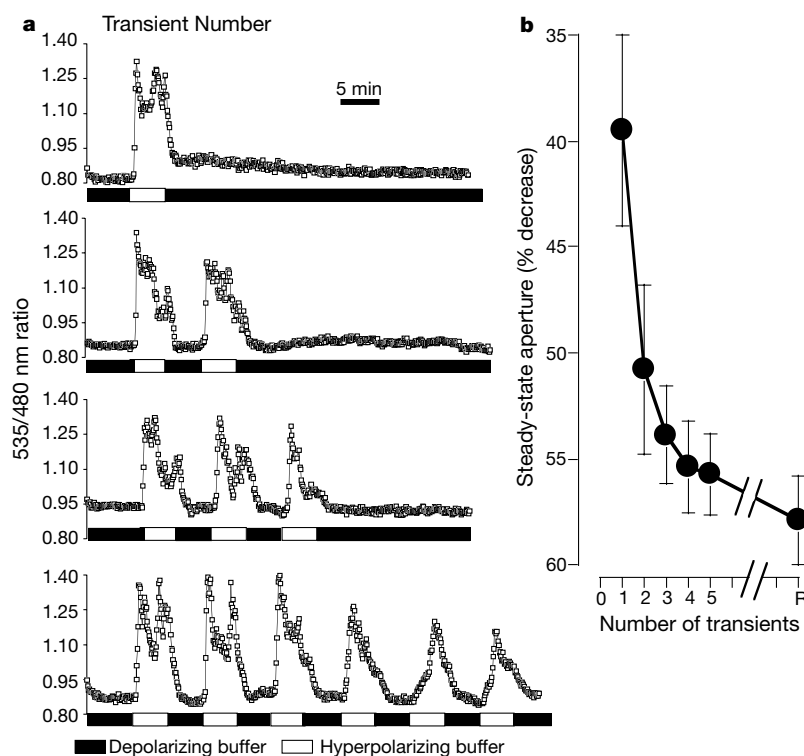


Figure 1 Increasing the number of $[Ca^{2+}]_{cyt}$ transients decreases steady-state stomatal aperture. **a**, Increasing number of $[Ca^{2+}]_{cyt}$ transients imposed in guard cells by repetitive exchanges between depolarizing and hyperpolarizing buffers. **b**, Steady-state stomatal aperture following increasing numbers of $[Ca^{2+}]_{cyt}$ transients. R denotes repetitive

exchanges inducing 6–9 transients (as in bottom panel of **a**). Steady-state aperture changes were calculated as a percentage decrease and are plotted as descending points to indicate stomatal closure.

measured 3 h after the start of the oscillation. An increasing number of $[Ca^{2+}]_{cyt}$ transients elicited progressively larger decreases in steady-state stomatal aperture, with four or more transients causing a decrease of 55–60% (Fig. 1b). Enhanced stomatal closing induced by increasing the number of $[Ca^{2+}]_{cyt}$ transients (Fig. 1a, b) was not due to guard cells responding to a greater overall $[Ca^{2+}]_{cyt}$ increase, because three 5-min $[Ca^{2+}]_{cyt}$ transients elicited much greater steady-state stomatal closure than a 15-min $[Ca^{2+}]_{cyt}$ plateau³ even though the total $[Ca^{2+}]_{cyt}$ increase was similar ($P > 0.56$; see Supplementary Information Fig. 2; total integrated $[Ca^{2+}]_{cyt}$ increase was $3,332 \pm 422 \text{ nM min}^{-1}$ for three 5-min transients ($n = 22$) and $3,678 \pm 97 \text{ nM min}^{-1}$ for 15-min plateau ($n = 30$)).

Three imposed $[Ca^{2+}]_{cyt}$ transients had similar magnitudes (Fig.

1a), showed little attenuation and elicited near-maximal stomatal closure ($54 \pm 2.2\%$, Fig. 1b), and were therefore used as a standard treatment to investigate whether other $[Ca^{2+}]_{cyt}$ oscillation parameters affected stomatal aperture. The effect of $[Ca^{2+}]_{cyt}$ oscillation frequency on steady-state stomatal aperture was investigated by imposing oscillations with latencies of 2, 5, 10 or 15 min between three 5-min $[Ca^{2+}]_{cyt}$ transients. The resulting oscillations had periods of 7, 10, 15 or 20 min (Fig. 2a). An optimal $[Ca^{2+}]_{cyt}$ oscillation period of 10 min was apparent (Fig. 2b) with reduced stomatal closure at longer or shorter periods. These data indicate that in guard cells an optimal frequency of $[Ca^{2+}]_{cyt}$ oscillation is required to elicit a maximal reduction in steady-state stomatal aperture.

By varying the low-KCl buffer perfusion time, the duration of

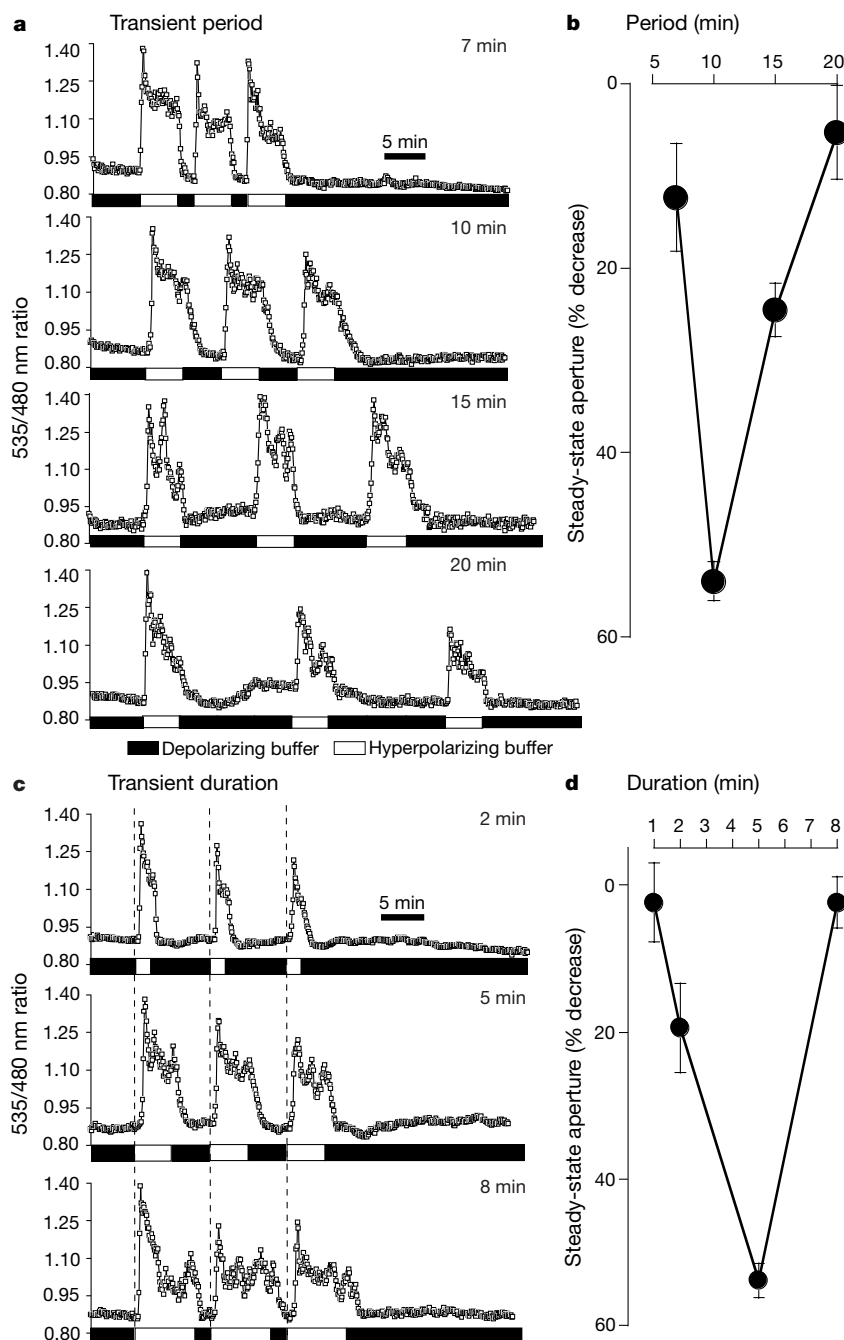


Figure 2 The period and duration of $[Ca^{2+}]_{cyt}$ transients alters steady-state stomatal aperture. **a**, Three 5-min $[Ca^{2+}]_{cyt}$ transients imposed in guard cells with 7, 10, 15 or 20-min periods. **b**, Steady-state stomatal aperture following imposed $[Ca^{2+}]_{cyt}$ oscillations as

in **a**. **c**, Three $[Ca^{2+}]_{cyt}$ transients imposed in guard cells with 2, 5 or 8-min duration. Dashed lines indicate the fixed period. **d**, Stomatal aperture following imposed $[Ca^{2+}]_{cyt}$ oscillations as in **c**. See also Supplementary Information Fig. 3.

$[Ca^{2+}]_{cyt}$ transients in guard cells was varied with a constant 10-min period. Three transients were imposed, each with a 1, 2, 5 or 8-min duration (Fig. 2c). Transients with a 5-min duration elicited the greatest decrease in steady-state stomatal aperture (Fig. 2d), whereas closure was reduced with 2 or 8-min transients (Fig. 2d). A similar optimum for stomatal closure was measured if the transient duration was varied and the high-KCl buffer perfusion time between $[Ca^{2+}]_{cyt}$ transients was kept constant at 5 min (data not shown). These data indicate that an optimal $[Ca^{2+}]_{cyt}$ transient duration during $[Ca^{2+}]_{cyt}$ oscillations is also required to elicit maximal stomatal closure (Fig. 2d). A minimum transient duration greater than 1 min was also required to elicit significant steady-state closure (Fig. 2d; see also Supplementary Information Fig. 3). Reducing the amplitude of $[Ca^{2+}]_{cyt}$ oscillations also reduced stomatal closure (see Supplementary Information Fig. 4).

To investigate short-term regulation of stomatal aperture during the progression of $[Ca^{2+}]_{cyt}$ oscillations, stomatal apertures were measured 1 min after each $[Ca^{2+}]_{cyt}$ transient during imposed oscillations with 10 or 20-min periods (Fig. 3). Oscillations with these periods gave the greatest difference in steady-state aperture change (Fig. 2b). Stomatal aperture decreased by around 45% immediately after an initial $[Ca^{2+}]_{cyt}$ transient and decreased further following subsequent transients by a maximum of around 65%, irrespective of the oscillation period (Fig. 3a, b). When the oscillation had a period of 10 min, steady-state stomatal apertures were still reduced by around 55% after 3 h (Fig. 3a). Following an oscillation with a 20-min period, however, stomata re-opened over the subsequent 1–1.5 h to reach a steady-state aperture that was close to that before the oscillation (Fig. 3b). A 20-min $[Ca^{2+}]_{cyt}$ plateau that failed to evoke a steady-state change in aperture³ (Fig. 3c) also induced short-term stomatal closure. Figure 3 shows that $[Ca^{2+}]_{cyt}$ oscillations cause stomatal closure by two mechanisms. The first is 'calcium reactive' and causes a rapid, short-term closure when $[Ca^{2+}]_{cyt}$ is elevated. The second is a long-term 'calcium programmed' decrease in the steady-state aperture, the degree of which is encoded by the frequency, transient number, duration and amplitude of $[Ca^{2+}]_{cyt}$ oscillations (Figs 1–3). Therefore, following calcium-reactive closure, guard cells can elicit programmed changes in aperture by decoding the information contained in $[Ca^{2+}]_{cyt}$ oscillation parameters. The appropriate $[Ca^{2+}]_{cyt}$ oscillation parameters inhibited stomatal opening.

We screened *abi1-1*, *abi2-1* (ref. 25), *eral-2* (ref. 26) and *gca2* (ref. 8) *Arabidopsis* mutants, all of which exhibit aberrant guard cell abscisic acid (ABA) signalling^{8,25,27,28}, to investigate whether these mutations altered $[Ca^{2+}]_{cyt}$ oscillation kinetics. We found that $[Ca^{2+}]_{ext}$ -induced $[Ca^{2+}]_{cyt}$ oscillation kinetics were markedly different from the wild type in *gca2* guard cells (Fig. 4). The recessive mutant *gca2* (*growth controlled by abscisic acid*) is insensitive to ABA in its root growth and stomatal response⁸. Stomata of *gca2* are also impaired in the closure response to H_2O_2 , and activation of a plasma membrane calcium channel by H_2O_2 is disrupted²⁴. In wild-type guard cells, endogenous $[Ca^{2+}]_{cyt}$ oscillations elicited by raising $[Ca^{2+}]_{ext}$ from 50 μ M to 10 mM (with constant 5-mM KCl buffer^{3,14,17,20}) had a period of 8.2 ± 1.7 min and a transient duration of 3.9 ± 0.6 min ($n = 46$ cells), whereas in *gca2* guard cells 10 mM $[Ca^{2+}]_{ext}$ elicited oscillations of similar amplitude ($P > 0.46$) but with a much shorter period (3.6 ± 0.7 min, $P < 0.001$) and transient duration (1.6 ± 0.2 min, $P < 0.001$; $n = 38$ cells; Fig. 4a). Steady-state stomatal closure in response to $[Ca^{2+}]_{ext}$ was abolished in *gca2* (Fig. 4b). The physiological stimulus ABA (5 μ M) elicited $[Ca^{2+}]_{cyt}$ oscillations with a period of 10.3 ± 0.7 min and a transient duration of 3.3 ± 0.3 min in wild-type guard cells ($n = 23$ cells; Fig. 4c). In *gca2* guard cells, ABA-induced oscillations had a significantly reduced period of 5.7 ± 0.5 min ($n = 24$, $P < 0.00003$) and transient duration of 1.9 ± 0.1 min ($P < 0.0004$; Fig. 4c). Stomatal closure in *gca2* was also insensitive to ABA (Fig. 4d).

To investigate whether Ca^{2+} reactive or programmed stomatal

closure is disrupted in the *gca2* mutant, $[Ca^{2+}]_{cyt}$ oscillations composed of three 5-min transients at 10-min periods were experimentally imposed in *gca2* guard cells ($n = 24$, Fig. 5a). Both calcium-reactive short-term closure and calcium-programmed long-term closure were restored to around 75% of the wild-type level by these

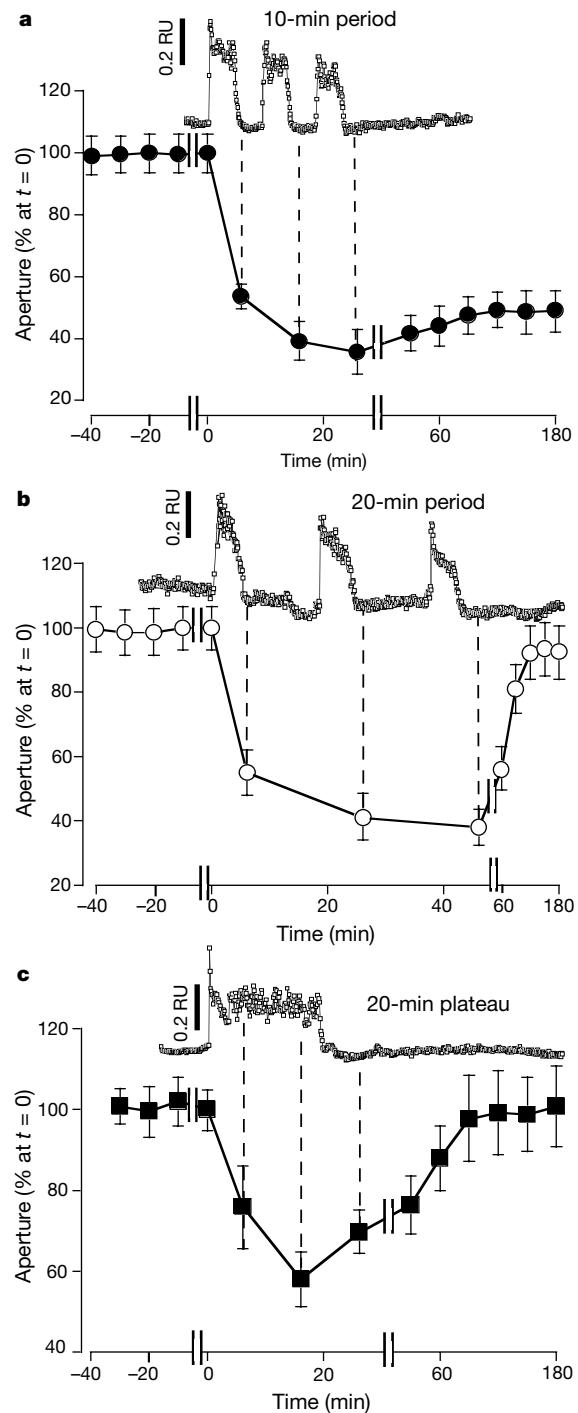


Figure 3 Preceding $[Ca^{2+}]_{cyt}$ oscillation kinetics program steady-state stomatal aperture. **a**, Stomatal aperture changes measured during imposed $[Ca^{2+}]_{cyt}$ oscillations with periods of 10 min. **b**, Stomatal aperture changes measured during imposed $[Ca^{2+}]_{cyt}$ oscillations with periods of 20 min. **c**, Stomatal aperture changes measured during 20-min imposed $[Ca^{2+}]_{cyt}$ plateau³. Apertures were measured for the last 40 min of the 2.5-h opening period (before time $t = 0$) and for a total of 3 h after $t = 0$. Cells were returned to the depolarizing buffer following the imposition of $[Ca^{2+}]_{cyt}$ increases. Note scale changes on x-axis are for aperture measurements only. RU = 535/480-nm ratio units. Data are the mean $\pm$ s.e.m. relative to $n = 4$ replicates comprising 12 individually mapped stomates.

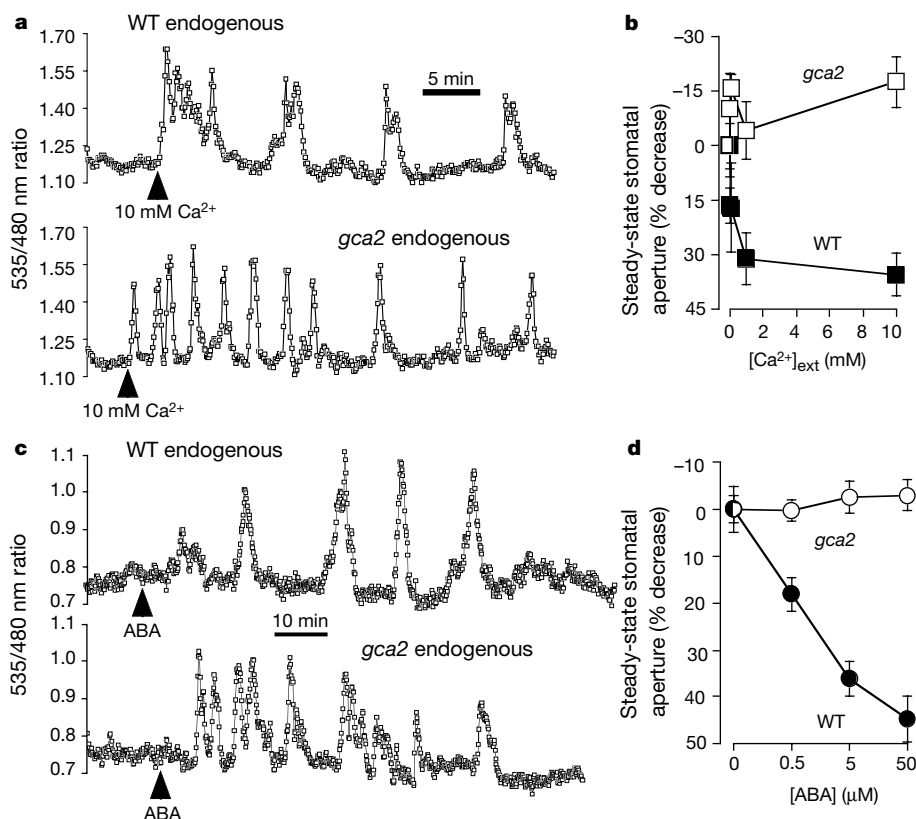


Figure 4 Altered [Ca^{2+}]_{cyt} oscillation kinetics and stomatal closure in the *gca2* mutant. **a**, [Ca^{2+}]_{cyt} oscillations elicited by 10 mM [Ca^{2+}]_{ext} in wild-type (top) and *gca2* (bottom) guard cells. **b**, Steady-state stomatal aperture changes elicited by [Ca^{2+}]_{ext} in wild-type (closed symbols) and *gca2* (open symbols). **c**, [Ca^{2+}]_{cyt} oscillations elicited by 5 μM ABA in

wild-type (top) and *gca2* (bottom) guard cells. **d**, Steady-state stomatal aperture changes elicited by ABA in wild-type (closed symbols) and *gca2* (open symbols). Data in **b** and **d** are mean $\pm$ s.e.m. relative to $n = 3$ replicates comprising 60 stomates per point.

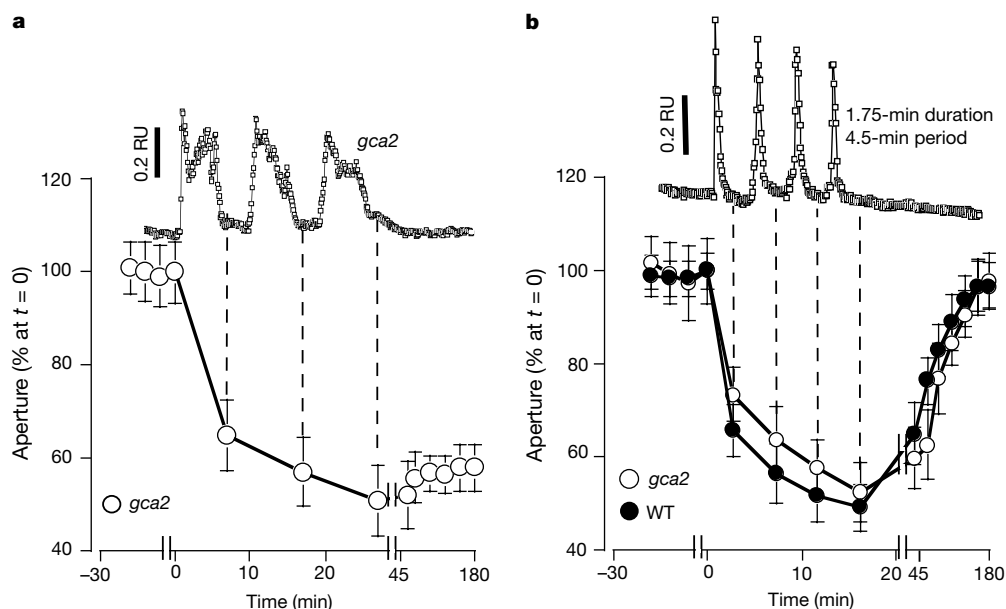


Figure 5 Altered [Ca^{2+}]_{cyt} oscillation kinetics in *gca2* prevent long-term stomatal closure. **a**, Stomatal aperture changes measured during imposed [Ca^{2+}]_{cyt} oscillations with periods of 10 min in *gca2* guard cells. **b**, Stomatal aperture changes measured in wild type and *gca2* during imposed [Ca^{2+}]_{cyt} oscillations that mimic stimulus-induced oscillations in *gca2* guard cells. Note imposed oscillations had an amplitude $\sim 30\%$ greater than

endogenous oscillations, which may enhance short-term responses. Cells were returned to the depolarizing buffer after the imposition of the oscillation. Scale changes on x-axis are for aperture measurements only. RU = 535/480-nm ratio units. Data are the mean $\pm$ s.e.m. relative to $n = 3$ replicates comprising nine individually mapped stomates per trace.

imposed oscillations (Fig. 5a, compare to Fig. 3a). The 75% response indicates that *gca2* may have partial effects on additional mechanisms downstream of $[Ca^{2+}]_{cyt}$ increases. Additionally, to mimic stimulus-induced $[Ca^{2+}]_{cyt}$ oscillations in *gca2* guard cells, oscillations composed of four $[Ca^{2+}]_{cyt}$ transients of 1.75 min duration at 4.5-min periods were imposed in both wild-type and *gca2* guard cells (Fig. 5b). These parameters represent the mean duration and frequency of $[Ca^{2+}]_{ext}$ and ABA-induced oscillations in *gca2* guard cells (Fig. 4a, c). These oscillations elicited calcium-reactive stomatal closure in both wild-type and *gca2* cells, but failed to elicit any long-term change in aperture through the calcium-programmed response (Fig. 5b). Overall, Fig. 5 shows that *gca2* guard cells remain competent to respond to $[Ca^{2+}]_{cyt}$ oscillations if the transient duration and frequency are within the range required to elicit long-term steady-state stomatal closure (Fig. 5a). Furthermore, stimulus-induced $[Ca^{2+}]_{cyt}$ oscillations in *gca2* guard cells, although sufficient to cause calcium-reactive closure, have kinetics outside the 'window' necessary to elicit a calcium-programmed change in steady-state aperture.

The $[Ca^{2+}]_{cyt}$ oscillations imposed in this study are similar to previously reported guard cell $[Ca^{2+}]_{cyt}$ oscillations induced by physiological stimuli: transients are asymmetric (the rise in $[Ca^{2+}]_{cyt}$ being faster than the decay) and successive transients show attenuation^{3,14,17,23,24,29,30}. ABA, oxidative stress and $[Ca^{2+}]_{ext}$ induce $[Ca^{2+}]_{cyt}$ oscillations in guard cells with periods between 6 and 12 min and induce 25–50% stomatal closure^{3,14,17,24,30}, which is consistent with the degree of stomatal closure induced by similar imposed oscillations in this study. Recently, $[Ca^{2+}]_{ext}$ was shown to elicit $[Ca^{2+}]_{cyt}$ plateaux in guard cells of the *Arabidopsis* V-ATPase mutant *det3* and steady-state stomatal closure was abolished³. However, the *det3* response gives no indication of whether individual $[Ca^{2+}]_{cyt}$ oscillation parameters affect stomatal closure. The alteration of the frequency and transient duration of $[Ca^{2+}]_{cyt}$ oscillations in *gca2* guard cells and the disruption of steady-state stomatal closure (Figs 4, 5) clearly show that the capacity to generate appropriate guard cell $[Ca^{2+}]_{cyt}$ oscillations is affected in *gca2*. This alteration of stimulus-induced $[Ca^{2+}]_{cyt}$ oscillation kinetics is the predominant factor in preventing a long-term calcium-programmed change in stomatal aperture. To our knowledge, disruption of calcium-programmed stomatal closure in *gca2* and the alteration of a periodic defecation pattern in the *dec-4* *Caenorhabditis elegans* mutant¹² are the only examples for which modulation of a physiological response has been directly correlated to a change in $[Ca^{2+}]_{cyt}$ oscillation kinetics. Overall, these data provide physiological and genetic evidence that the kinetics of $[Ca^{2+}]_{cyt}$ oscillations in guard cells can control steady-state stomatal aperture changes^{4,18} and therefore show that $[Ca^{2+}]_{cyt}$ oscillation kinetics in plant cells can encode information that controls the magnitude of a graded physiological response. □

Methods

Plant growth

A. thaliana was grown with a 16-h light/8-h dark cycle and a photon fluency rate of $75 \mu\text{mol m}^{-2} \text{s}^{-1}$ at 20 °C.

Calcium imaging

We performed calcium imaging and calibration as described^{3,20}. We incubated abaxial epidermal strips for 2.5 h in the light ($125 \mu\text{mol m}^{-2} \text{s}^{-1}$) in 100 mM KCl, 0 CaCl₂ and 10 mM Mes-Tris pH 6.15 (depolarizing buffer) to open stomata. $[Ca^{2+}]_{cyt}$ transients were imposed by rapid (10 s) exchange of the chamber buffer between this depolarizing buffer and 0.1 mM KCl, 10 mM CaCl₂ and 10 mM Mes-Tris pH 6.15 (hyperpolarizing buffer). We measured imposed oscillations in 16–38 cells for each treatment; they were highly reproducible between cells. Baseline decreases in emission ratios due to bleaching were subtracted. The 535/480-nm ratios are smaller than previously reported^{13,20,24} because we used a CCD camera with improved 440-nm sensitivity. We measured endogenous oscillations (Fig. 4) in 5 mM KCl, 50 μM CaCl₂ and 10 mM Mes-Tris pH 5.6. In these conditions, 32% of guard cells exhibited spontaneous oscillations³⁰, which were similar in wild-type and *gca2* ($P > 0.5$). Only cells exhibiting stable resting $[Ca^{2+}]_{cyt}$ were used to analyse stimulus-induced oscillations. ABA-induced $[Ca^{2+}]_{cyt}$ increases were observed in 72% of cells (47 of 65 cells).

Stomatal aperture measurements

Stomata were opened and oscillations imposed using the same method as for imaging. After imposition of an oscillation, cells were maintained in the (100 mM KCl) depolarizing buffer and steady-state changes in aperture ratio (width/length) were measured as described³ before and 3 h after the initiation of $[Ca^{2+}]_{cyt}$ oscillations (Figs 1, 2 and 4) or throughout the oscillation (Figs 3 & 5). Apertures are mean $\pm$ s.e.m. relative to $n = 4$ replicates comprising 160 stomata, unless otherwise stated.

Received 27 October 2000; accepted 27 April 2001.

- Berridge, M. J. The AM and FM of calcium signalling. *Nature* **386**, 759–760 (1997).
- MacRobbie, E. A. C. Signalling in guard cells and regulation of ion channel activity. *J. Exp. Bot.* **48**, 515–528 (1997).
- Allen, G. J. et al. Alteration of stimulus-specific guard cell calcium oscillations and stomatal closing in *Arabidopsis det3* mutant. *Science* **289**, 2338–2342 (2000).
- McAinsh, M. R. & Hetherington, A. M. Encoding specificity in Ca^{2+} signalling systems. *Trends Plant Sci.* **3**, 32–36 (1998).
- Gu, X. & Spitzer, N. C. Distinct aspects of neuronal differentiation encoded by frequency of spontaneous Ca^{2+} transients. *Nature* **375**, 784–787 (1995).
- Dolmetsch, R. E., Xu, K. & Lewis, R. S. Calcium oscillations increase the efficiency and specificity of gene expression. *Nature* **392**, 933–936 (1998).
- De Koninck, P. & Schulman, H. Sensitivity of CaM kinase II to the frequency of Ca^{2+} oscillations. *Science* **279**, 227–230 (1998).
- Himmelbach, A., Iten, M. & Grill, E. Signalling of abscisic acid to regulate plant growth. *Phil. Trans. R. Soc. Lond. B* **353**, 1439–1444 (1998).
- Li, W. et al. Cell-permeant cage InsP_3 ester shows that Ca^{2+} spike frequency can optimize gene expression. *Nature* **392**, 936–941 (1998).
- Gomez, T. M. & Spitzer, N. C. *In vivo* regulation of axon extension and path finding by growth-cone calcium transients. *Nature* **397**, 350–355 (1999).
- Camacho P. & Lechleiter, J. D. Increased frequency of calcium waves in *Xenopus laevis* oocytes that express a calcium-ATPase. *Science* **260**, 226–229 (1993).
- Dal Santo, P., Logan, M. A., Chisholm, A. D. & Jorgensen, E. M. The inositol trisphosphate receptor regulates a 50-second behavioral rhythm in *C. elegans*. *Cell* **98**, 757–767 (1999).
- Felle, H. Auxin causes oscillations of cytosolic free calcium and pH in *Zea mays* coleoptiles. *Planta* **174**, 495–499 (1988).
- McAinsh, M. R., Webb, A. A. R., Taylor, J. E. & Hetherington, A. M. Stimulus-induced oscillations in guard cell cytoplasmic free calcium. *Plant Cell* **7**, 1207–1219 (1995).
- Ehrhardt, D. W., Wais, R. & Long, S. R. Calcium spiking in plant root hairs responding to rhizobium nodulation signals. *Cell* **85**, 673–681 (1996).
- Holdaway-Clarke, T. L., Feijó, J. A., Hackett, G. R., Kunkel, J. G. & Hepler, P. K. Pollen tube growth and intracellular cytosolic calcium gradient oscillate in phase while extracellular calcium influx is delayed. *Plant Cell* **9**, 1999–2010 (1997).
- Staxén, I. et al. Abscisic acid induces oscillations in guard-cell cytosolic free calcium that involve phosphoinositide-specific phospholipase C. *Proc. Natl Acad. Sci. USA* **96**, 1779–1784 (1999).
- Trewavas, A. J. & Malho, R. Ca^{2+} signalling in plant cells: the big network! *Curr. Opin. Plant Biol.* **1**, 428–433 (1998).
- Miyawaki, A., Griesbeck, O., Heim, R. & Tsien, R. Y. Dynamic and quantitative Ca^{2+} measurements using improved cameleons. *Proc. Natl Acad. Sci. USA* **96**, 2135–2140 (1999).
- Allen, G. J. et al. Cameleon calcium indicator reports cytoplasmic calcium dynamics in *Arabidopsis* guard cells. *Plant J.* **19**, 735–747 (1999).
- Gilroy, S., Fricker, M. D., Read, N. D. & Trewavas, A. J. Role of calcium in signal transduction of commelina guard-cells. *Plant Cell* **3**, 333–344 (1991).
- Grabov, A. & Blatt, M. R. Membrane voltage initiates Ca^{2+} waves and potentiates Ca^{2+} increases with abscisic acid in stomatal guard cells. *Proc. Natl Acad. Sci. USA* **95**, 4778–4783 (1998).
- Hamilton, D. W., Hills, A., Kohler, B. & Blatt, M. R. Ca^{2+} channels at the plasma membrane of stomatal guard cells are activated by hyperpolarization and abscisic acid. *Proc. Natl Acad. Sci. USA* **97**, 4967–4972 (2000).
- Pei, Z. M. et al. Hydrogen peroxide-activated Ca^{2+} channels mediate guard cell abscisic acid signalling. *Nature* **406**, 731–734 (2000).
- Koornneef, M., Reuling, G. & Karssen, C. M. The isolation and characterisation of abscisic acid-insensitive mutants of *Arabidopsis thaliana*. *Physiol. Plant* **61**, 377–383 (1984).
- Cutler, S., Ghassemian, M., Bonetta, D., Cooney, S. & McCourt, P. A protein farnesyl transferase involved in abscisic acid signal transduction in *Arabidopsis*. *Science* **273**, 1239–1241 (1996).
- Pei, Z. M., Ghassemian, M., Kwak, C. M., McCourt, P. & Schroeder, J. I. Role of farnesyltransferase in ABA regulation of guard cell anion channels and plant water loss. *Science* **282**, 287–290 (1998).
- Pei, Z. M. et al. Differential abscisic acid regulation of guard cell slow anion channels in *Arabidopsis* wild-type and *abi1* and *abi2* mutants. *Plant Cell* **9**, 409–423 (1997).
- Schroeder, J. I. & Hagiwara, S. Repetitive increases in cytoplasmic Ca^{2+} of guard cells by abscisic acid activation of non-selective Ca^{2+} -permeable channels. *Proc. Natl Acad. Sci. USA* **87**, 9305–9309 (1990).
- Blatt, M. R. Ca^{2+} signalling and control of guard-cell volume in stomatal movements. *Curr. Opin. Plant Biol.* **3**, 196–204 (2000).

Supplementary Information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank J. Harper for discussions and N. Spitzer and R. Tsien for comments on the manuscript. This research was supported by NIH, Department of Energy and National Science Foundation grants to J.I.S. and a DFG grant and Fonds der Chemischen Industrie to E.G.

Correspondence and requests for materials should be addressed to J.I.S. (e-mail: Julian@biomail.ucsd.edu).

Roles of tumour localization, second signals and cross priming in cytotoxic T-cell induction

Adrian F. Ochsenbein^{*†}, Sophie Sierro^{*}, Bernhard Odermatt^{*}, Marcus Pericin[†], Urs Karrer^{*}, Jan Hermans[‡], Silvio Hemmi[§], Hans Hengartner^{*} & Rolf M. Zinkernagel^{*}

^{*} Institute of Experimental Immunology, University Hospital, CH-8091 Zurich, Switzerland

[‡] Malaghan Institute of Medical Research, Wellington South, New Zealand

[§] Institute of Molecular Biology, University of Zurich, CH-8057 Zurich

The vertebrate immune system has evolved to protect against infections that threaten survival before reproduction. Clinically manifest tumours mostly arise after the reproductive years and somatic mutations allow even otherwise antigenic tumours to evade the attention of the immune system^{1–3}. Moreover, the lack of immunological co-stimulatory molecules on solid tumours could result in T-cell tolerance^{4–8}; that is, the failure of T cells to respond. However, this may not generally apply^{9,10}. Here we report several important findings regarding the immune response to tumours, on the basis of studies of several tumour types. First, tumour-specific induction of protective cytotoxic T cells (CTLs) depends on sufficient tumour cells reaching secondary lymphatic organs early and for a long enough duration. Second, diffusely invading systemic tumours delete CTLs. Third, tumours that stay strictly outside secondary lymphatic organs, or that are within these organs but separated from T cells by barriers, are ignored by T cells but do not delete them. Fourth, co-stimulatory molecules on tumour cells do not influence CTL priming but enhance primed CTL responses in peripheral solid tumours. Last, cross priming of CTLs by tumour antigens, mediated by major histocompatibility complex (MHC) class I molecules of antigen-presenting host cells, is inefficient and not protective. These rules of T-cell induction and maintenance not only change previous views but also rationales for anti-tumour immunotherapy¹².

Tumours may escape the attention of the immune system and metastasize through various mechanisms, including loss of tumour or MHC antigens and/or the expression of inhibitory molecules^{1,2}. However, even antigenic tumours can grow successfully and may persist in a host with an intact, fully functioning immune system^{2,3}. This is reminiscent of immunity against pathogens where peripheral, low-level infections coexist with T-cell immunity, resulting in the immune system being incapable of clearing all peripheral infected cells¹¹. Failure of immunity against tumours has been explained by deletion of T cells that interact with tumour cells lacking co-stimulatory signals^{4–8}. However, a fibrosarcoma was shown not to delete CTLs but rather to prime them directly if tumour cells reached secondary lymphoid tissues early^{9,10}. These results were termed as an exception and questioned because antigen-presenting cells (APCs) of the host may have reprocessed tumour antigen onto MHC class I molecules (termed 'cross priming')^{12–16}. Such cross priming has been observed *in vitro*, and occasionally also at the peptide level *in vivo*^{12,13,17}. The relative role of CTLs versus CD4⁺ T-helper cells in cross priming *in vivo* remains unclear^{17,18}.

We studied MC57G (C57BL/6 strain, haplotype H-2^b) and D2 (B10.D2, H-2^d) fibrosarcoma and EL4 (C57BL/6, H-2^b) lymphoma cells transfected with the entire lymphocytic choriomeningitis virus glycoprotein (LCMV-GP) as tumour anti-

gens (MC-GP, D2-GP and EL4-GP)^{9,10}, and also the Lewis lung carcinoma (LL) (C57BL/6, H-2^b) (LL-GP33)¹⁹, EL4 lymphoma (EL4-GP33) and B16.F10 melanoma (C57BL/6, H-2^b) (B16-GP33)²⁰ cell lines transfected with the GP33 D^b epitope. P815 mastocytoma cells (DBA/2, H-2^d) were transfected with the entire LCMV nucleoprotein (P815-NP). MC-GP, LL-GP33 and EL4-GP33 cells expressed high levels of MHC class-I D^b, whereas B16-GP33 cells expressed lower levels (Fig. 1A, a). Incubation in 10 U ml⁻¹ interferon- γ (IFN- γ) for 48 h increased H-2D^b expression 100-fold (Fig. 1A, c). MC-GP and LL-GP33 cells did not express B7.1, B7.2, lymphocyte function-associated antigen (LFA-1) or intercellular adhesion molecule-1 (ICAM-1) above background levels. B16-GP33 cells expressed ICAM-1 at very low levels (not shown) and EL4-GP33 exhibited B7.1, LFA-1 and ICAM-1, but not B7.2 (Fig. 1A, b). CTL-target-cell qualities of all cell lines were analysed (Fig. 1A, d); they confirmed earlier experiments^{10,19,20}.

We first established that strictly peripheral, extralymphatic tumours are ignored by the immune system. We injected single-cell suspensions or small tumour fragments of 2–5 × 10⁶ tumour cells subcutaneously¹⁰. We then compared tumour growth, CTL induction and migration of the transferred tumour cells to secondary lymphoid organs. Eleven of 20 EL4-GP33 (Fig. 1B, a) pieces grew. All transplanted tumour pieces grew for LL-GP33 (Fig. 1B, e) and B16-GP33; 15 of 20 MC-GP pieces grew (not shown but see ref. 10). Without exception tumour growth correlated with absence of primed CTLs (Fig. 1B, b, d and f). Absence of viable (or polymerase chain reaction (PCR)-detectable¹⁰) tumour cells in secondary lymphoid organs (Table 1) correlated with absence of primed CTLs (Fig. 1B, b, d and f). In contrast, rejection of EL4-GP33 and MC-GP tumour pieces paralleled induction of GP33-specific CTLs (Fig. 1B, b). Distinct from tumour fragments, MC-GP and EL4-GP33 cell suspensions injected subcutaneously in titrated doses from 10⁴ to 10⁸ cells were all rejected, but they did prime CTLs (Fig. 1B, a and b). In mice depleted of natural killer and CD8⁺ T cells, MC-GP and EL4-GP33 tumour cells were detectable by selective culture (or by PCR¹⁰) in local lymph nodes after injection of cells, but not of tumour fragments (Table 1). Rejection of 50% of EL4-GP33 tumours (Fig. 1B, a) correlated with the fragility of tumour fragments and cells reaching lymphoid organs at variable times: in mice depleted of CD8⁺ and natural killer cells, EL4-GP33 cells were found in lymph nodes in 1 of 4 mice on day 12 and in 3 of 4 mice on day 16 after transplantation. Notably, 9 of 14 transplanted MC-GP tumour pieces grew in minor histo-incompatible C57BL/10 mice (see Supplementary Information Fig. 1) and 2 of 4 grew in BALB/b mice. Thus, cells that are strictly peripheral failed to induce GP33-specific CTLs even when minor histo-incompatibility antigens provided additional CTL-epitopes and could induce CD4⁺ T-helper cells.

B16-GP33 and LL-GP33 cells also followed the pattern that

Table 1 Homing of tumour cells to secondary lymphoid organs

Tumour cells or fragments*	Detection of tumour cells†			
	Day 4		Day 8	
	Lymph node	Spleen	Lymph node	Spleen
MC-GP fragment	0/4	0/4	0/4	0/4
MC-GP cells	2/4	3/4	3/4	4/4
EL4-GP33 fragment	0/8	0/8	0/8	0/8
EL4-GP33 cells	0/8	1/4	6/6	4/4
3LL-GP33 fragment	0/4	0/4	0/4	0/4
3LL-GP33 cells	0/8	0/4	0/8	0/4
B16-GP33 fragment	0/4	0/4	0/4	0/4
B16-GP33 cells	0/8	0/4	0/8	0/4

* Cells (2 × 10⁶) or solid tumour fragments (containing 2–5 × 10⁶ tumour cells) were transplanted to both flanks.

† To avoid immediate rejection mice were CD8⁺ T-cell-depleted by treatment with 200 μ l anti-CD8 antibody (YTS169.4.2) i.p. 3 or 1 days¹⁰ before cell or fragment transfer, and depleted of NK cells by 30 μ l of anti-asialo GM1 antibody (Wako Biochemicals) i.v. on 1 day before transfer. Lymph nodes and spleen cell suspensions were cultured in selection medium (G-418). Values indicate positive detection in number of mice over total number tested.

[†] Present address: Institute of Medical Oncology, Inselspital, 2010 Berne, Switzerland.

tumour growth correlated with absence of primed CTLs; $\geq 10^4$ single cells of B16-GP33 and LL-GP33 grew subcutaneously in all animals, as found earlier^{19,20} (Fig. 1B, c and e). No viable tumour cells were detected in draining lymph nodes 4–8 days after injection (Table 1) and no CTLs were induced (Fig. 1B, d and f). Growing tumours were not CTL-escape mutants as they expressed D^b-GP33 as targets and were susceptible to CTLs (not shown).

The correlation between tumour growth and, at early time points, absence of viable tumour cells in lymphoid organs and the absence of an anti-tumour CTL response is not called into question by the rejection of a small number of tumour pieces. Sometimes few but sufficient tumour cells may reach draining lymph nodes later, as shown for EL4-GP33. Although this is difficult to quantify, about 10^3 to 10^4 cells (MC-GP, EL4-GP33) seem necessary (Fig. 1D, b). This conclusion is strengthened by experiments with mice lacking all lymph nodes (see below). Notably, in all tumour models, tumour-specific CTLs were neither anergized nor deleted in

tumour-bearing mice, as immunization with LCMV (or antigenic dendritic or tumour cells¹⁰) promptly primed CTLs (Fig. 1B, b, d and f).

CTLs were primed by 1,000 or more live MC-GP and EL4-GP33 (but not by up to 10^6 B16-GP33 and LL-GP33) tumour cells injected directly into the spleen (Fig. 1C, a and b versus c and d, respectively). Instead, B16-GP33 and LL-GP33 cells grew as solid tumours up to a diameter of 2 cm (see also Fig. 2). When tumour cells labelled with fluorescent dye (CFSE) were analysed immunohistochemically during the first four days, EL4-GP33 were diffusely distributed throughout the spleen as single cells (not shown). MC-GP cells were intermixed with CD8⁺ T cells (Fig. 2a, d, g). In contrast, B16-GP33 (Fig. 2b, e, h) and LL-GP33 tumours (Fig. 2c, f, i) grew solidly with only minimal CTL infiltration. Of note, the LL-GP33 and B16-GP33 cells were surrounded by haemostasis factors (factor VIII-related antigen surrounding B16-GP33 (Fig. 2k) or collagen (surrounding B16-GP33 (Fig. 2n) and LL-GP33 (Fig. 2o)) when analysed 14 days

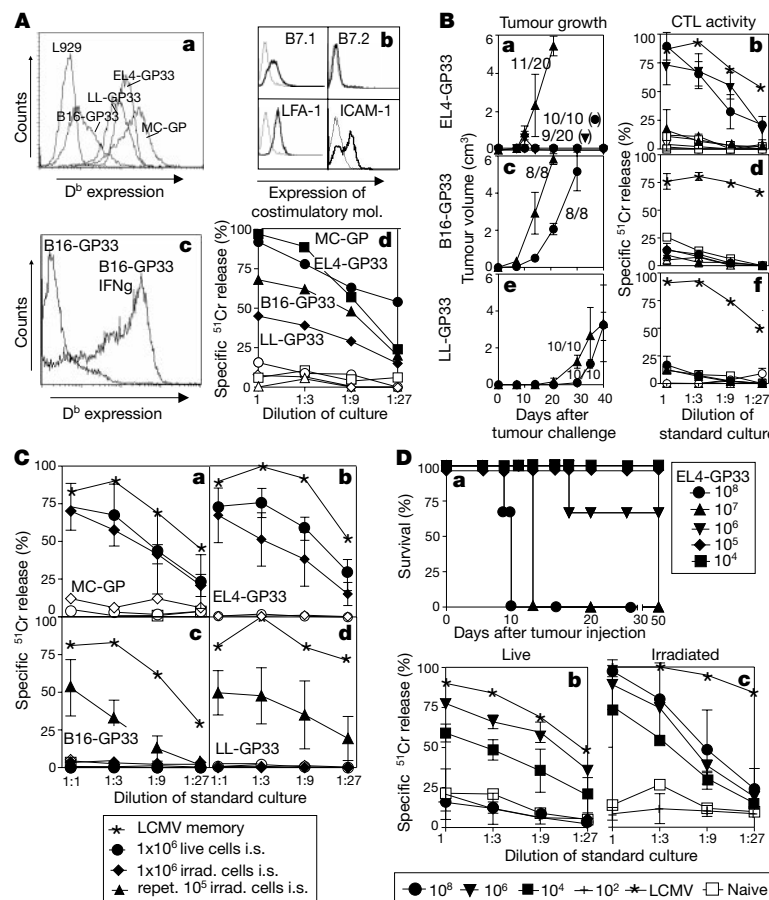


Figure 1 Ignorance, priming and tolerance of tumour-specific T cells. **A**, Characterization of the tumour cell lines. **a–d**, D^b expression (**a**), expression of second signals or adhesion markers on EL4-GP33 cells (**b**), D^b expression on B16-GP33 cells after 10 U IFN- γ for 48 h (**c**), lysis of MC-GP, EL4-GP33, LL-GP33 and B16-GP33 cells by LCMV-GP33-specific CTLs (**d**). Open symbols, untransfected control tumour cells. **B**, Comparison of tumour take versus induction of anti-tumour CTLs. **a–f**, EL4-GP33 (**a, b**), B16-GP33 (**c, d**) and LL-GP33 (**e, f**) single cells injected subcutaneously (filled circles) or implanted as small tumour pieces (filled triangles). Tumour growth (**a, c, e**) and secondary CTL activity (**b, d, f**) in mice with growing (filled, normal triangle) or rejected (filled, inverted triangle) tumours (numbers indicate growing tumours over the total number of transferred tumours). Secondary CTL activity against LCMV-GP33 targets after 5 days *in vitro* re-stimulation of spleen cells taken 8 days after injection of the tumour-cell suspension using unpulsed (open symbols) or GP33-pulsed (filled symbols) EL-4 (**d, f**) or MC57 (**b**), to avoid cross reactivities on EL4) as target cells. For tumour fragment experiments, CTL activity

was similarly assessed at the end of the experiment (after ~20 days for EL-4-GP33 and B16-GP33 (**b, d**) and 40 days for LL-GP33 (**f**)). CTL activity is given for each group as mean $\pm$ s.d. Some tumour-bearing mice were immunized with LCMV and 8 days later cytotoxicity of GP33 specific CTLs was assessed (asterisks). **C** Secondary CTL responses induced by 10^6 live tumour cells. **a–d**, MC-GP (**a**), EL4-GP33 (**b**), B16-GP33 (**c**) and LL-GP33 (**d**) cells injected directly into the spleen once (circles) and by single (diamonds) or repetitive (3 times on alternating days, CTL analysis 5 days after the last injection) (triangles) injection of irradiated tumour cells (MC-GP, 80 Gy; EL4-GP33, 40 Gy; B16-GP33, 200 Gy; LL-GP33, 100 Gy. LCMV-primed mice (200 PFU LCMV day 60) were used as controls (asterisks). **D**, Survival of C57BL/6 mice given EL4-GP33 lymphoma intravenously (**a**). Titrated numbers of 10^2 to 10^8 live (**b**) or irradiated (40 Gy); (**c**) EL4-GP33 cells were injected intravenously. Eight days later the secondary GP33-specific CTLs were assessed (mean $\pm$ s.d. of 3 to 4 mice per group). All experiments were done twice with similar results (see text).

after intrasplenic injection. In the absence of primed CTLs this suggested that naïve T cells were separated from antigen-expressing tumour cells and therefore were not induced. IFN γ -treated B16-GP33 cells expressing high levels of the D^b epitope (Fig. 1A, c) formed comparably separated tumours (as in Fig. 2h, k, n) and also failed to prime CTLs after splenic injection of 2×10^6 cells. Thus, low immunogenicity of B16-GP33 cells did not correlate with low expression of MHC class I but rather with barrier formation. When these tumour-bearing mice were immunized with 200

plaque-forming units (PFU) of LCMV by intraperitoneal injection, activated CTLs massively infiltrated splenic tumours 6 days later (Fig. 2p, q; MC-GP tumours were rejected at this time and cannot be shown), thereby negating suppression or tumour inaccessibility explanations.

To avoid cell proliferation and formation of barriers, irradiated cells (MC-GP, 80 Gy; EL4-GP33, 40 Gy; B16-GP33, 200 Gy; LL-GP33, 100 Gy; determined minimal irradiation doses) were injected directly into the spleen. A single injection of $\geq 10^4$ irradiated MC-

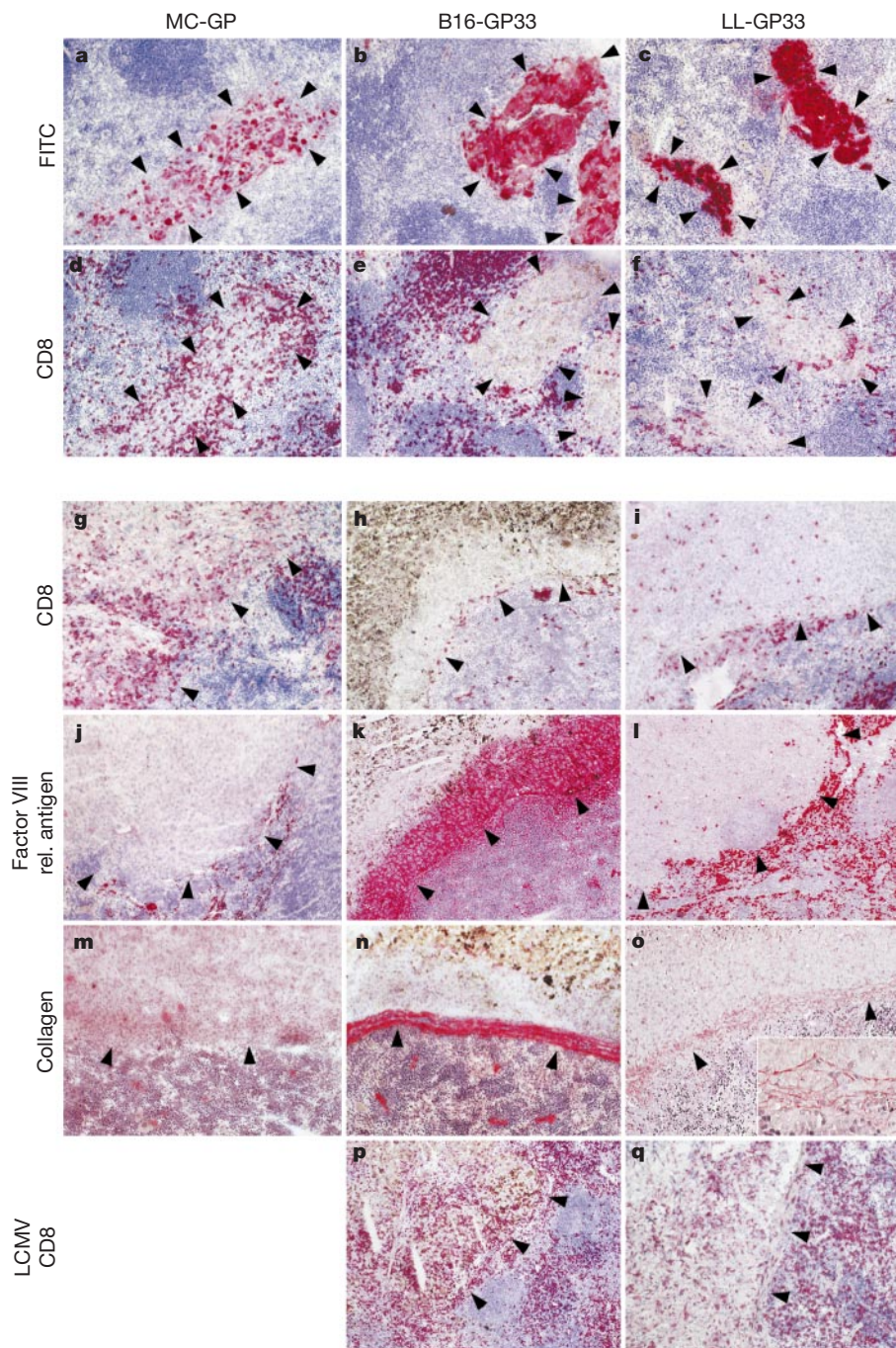


Figure 2 Analysis of tumour growth in lymphoid organs. CFSE-labelled MC-GP, EL4-GP33, B16-GP33 and LL-GP33 cells (10^6) were injected directly into the spleen. **a–c**, Anit-fluorescein isothiocyanate (FITC) staining of tumour cells 48 h after injection. **d–f**, Distribution of CD8⁺ T cells assessed by anti-CD8⁺ staining. **g–q**, Fourteen days after injection of B16-GP33 (**h, k, n, p**) and LL-GP33 (**i, l, o, q**), and 8 days after injection

of MC-GP (**g, j, m**) cells, spleens were stained for CD8⁺ T cells (**g–i**), factor VIII-related antigen (**j–l**) or collagen (**m–o**). Eight days after intrasplenic injection of B16-GP33 (**p**) or LL-GP33 (**q**) tumour cells the recipient mice were infected intravenously with 200 PFU LCMV. After 6 days, the resulting immune response was analysed by anti-CD8⁺ T-cell staining. One out of three comparable experiments is shown.

GP or EL4-GP33 cells primed CTLs (Fig. 1C, a, b and D, c). In contrast, one dose of 10^6 irradiated B16-GP33 or LL-GP33 cells alone did not prime CTLs (Fig. 1C, c and d). However, three repetitive intrasplenic injections of 10^5 irradiated tumour cells on alternating days induced GP33-specific CTLs (Fig. 1C, c and d). Thus, compared with MC-GP or EL4-GP33 cells, about 10 times more B16-GP33 or LL-GP33 tumour cells were necessary to induce CTLs, perhaps owing to lower antigen or MHC class I expression (Fig. 1A, a and d). Importantly, 10–1,000 times fewer irradiated, and therefore short-lived, tumour cells (B16-GP33 or LL-GP33) primed naïve CTLs in spleens; whereas huge numbers of living and multiplying but barrier-separated cells in, or tumours strictly outside of, lymphatic organs failed to induce CTLs.

These relatively low numbers ($\geq 10^4$) of CTL-accessible tumour cells in secondary lymphoid organs during minimal periods (4–6 days) seem necessary and sufficient for CTL-induction. The data also indicate that perhaps some widely used tumours (such as B16 and LL multiply passaged *in vivo*) have been selected to evade immunosurveillance by rapidly inducing barriers that isolate them from T cells. Such mechanisms may be one basis of metastatic tumour growth in lymphoid tissues.

Immunosurveillance of tumours spreading diffusely within lymphoid tissues was studied using T-cell lymphomas EL4-GP33 and mastocytomas P815-NP injected intravenously. Attempts to intravenously inject $\geq 10^6$ MC-GP or LL-GP33 cells failed owing to embolisms. After an intravenous injection of $> 10^6$ EL4-GP33 cells, hosts died within 9–18 days (Fig. 1D, a). About 10^4 to 10^6 live or $\geq$

Table 2 Anti-LCMV protection in mice primed with 10^7 or 10^8 EL4-GP33 cells

Immunization	Cell dose	LCMV titre ($\log_{10}$) [*]	
		Spleen	Liver
–	10^5	3.2 ± 0.2 †	<1.7 †§
–	10^7	4.6 ± 0.4 NS	3.7 ± 0.8 NS
200 PFU LCMV (0)	10^7	<1.7 §	<1.7 §
200 PFU LCMV (5)	10^7	<1.7 §	<1.7 §
2×10^6 Vacc-YN4 (14)	10^7	<1.7 §	<1.7 §
–	0	5.0 ± 0.5	3.6 ± 0.2
200 PFU LCMV (14)	0	<1.7 §	<1.7 §

Numbers in parentheses indicate the day of immunization before tumour challenge. NS, not significant.

^{*}Mice were infected 12 days after tumour injection with 10^4 PFU LCMV *i.v.* and LCMV titres per organ were assessed 4 days later. Results are given as mean $\pm$ s.d. of 3–5 animals per group. The experiment was repeated twice with comparable results.

† LCMV-WE replicates to higher titres in spleen than in liver. Overall 100-fold titre reduction by EL4-GP33-priming caused titres to drop in the liver below detection level, whereas in the spleen LCMV was still detectable in a plaque-forming assay.

‡ Students *t*-test (versus untreated mice ||); $P < 0.5$.

§ Students *t*-test (versus untreated mice ||); $P < 0.001$.

|| Untreated mice infected with 10^4 PFU LCMV.

10^4 irradiated EL4-GP33 cells (40 Gy) primed GP33-specific CTLs (Fig. 1D, b and c). Injection of high doses ($\geq 10^7$) of irradiated, but not of live, EL4-GP33 tumour cells induced CTLs (Fig. 1D, b and c). Similarly, $\geq 10^4$ irradiated P815-NP cells primed CTLs in DBA/2 mice, whereas $> 10^6$ live P815-NP cells induced CTL unresponsiveness as assessed by *in vitro* re-stimulation (not shown). However, excessive tumours in lymphoid organs might mask CTL induction

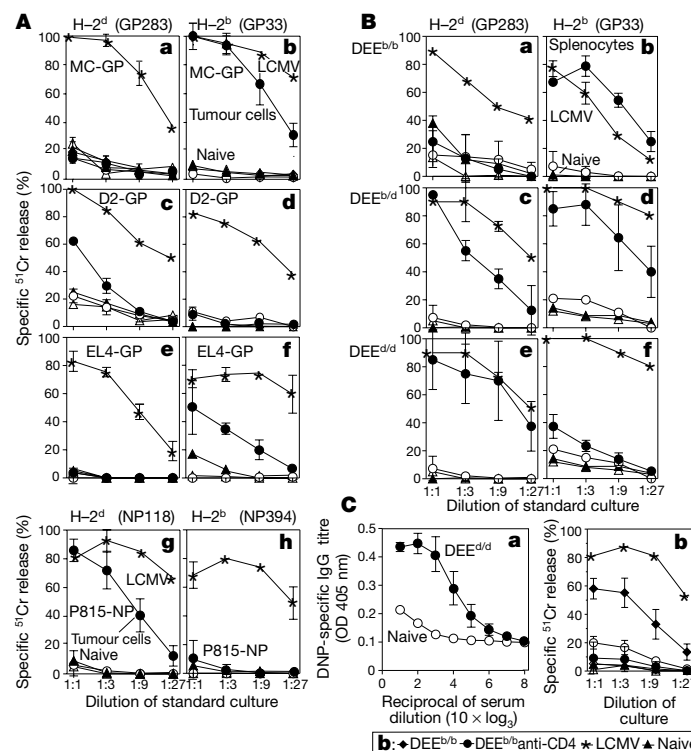


Figure 3 Antigen presentation for an anti-tumour immune response. **A**, CTL induction by 5×10^6 MC-GP (**a**, **b**), D2-GP (**c**, **d**) or EL4-GP cells (**e**, **f**) injected intraperitoneally into CB6 (C57BL/6 H-2^b $\times$ BALB/c H-2^d) F₁ animals. On day 14 we carried out re-stimulation *in vitro* on GP283-labelled (H-2^d) or on GP33-labelled (H-2^b) F₁ splenocytes for 5 days: cytotoxicity assay against GP283-labelled P815 (H-2^d) (**a**, **c**, **e**) or GP33-labelled EL4 (H-2^b) (**b**, **d**, **f**) targets. Using the same protocol 5×10^6 P815-NP cells (**g**, **h**) were injected into CB6 (C57BL/6 H-2^b $\times$ BALB/c H-2^d) F₁ mice: *in vitro* re-stimulation with peptide on F₁ cells and Cr-release assay against NP118-labelled (H-2d; **g**) or NP394-labelled (H-2b; **h**) target cells. **B**, CTL induction by 2×10^6 transgenic LCMV-GP-expressing splenocytes of DEE^{b/b} (**a**, **b**), DEE^{d/d} (**c**, **d**) or DEE^{d/d} (**e**, **f**) mice 14 days after

intravenous injection into CB6 mice. Re-stimulation and ^{51}Cr release assay as in **A**. **C**, T-helper cell induction. **a**, CB6 (H-2^d $\times$ H-2^b) mice treated with 2×10^6 DEE^{d/d} splenocytes intravenously and 12 days later challenged with 20 μg of DNP coupled to P13 (helper epitope GP60–80 in H-2^b (ref. 23)). Anti-DNP IgG titres were measured by ELISA on day 7. LCMV-primed (60 days, 10^2 PFU LCMV intravenously) mice served as positive controls. **b**, C57BL/6 mice either depleted of CD4⁺ T cells (filled circle) (day 3 and day 1, 200 μl anti-CD4⁺ antibody intraperitoneal injection) or left untreated (filled diamond). Mice were then treated with 2×10^6 DEE^{b/b} cells injected intraperitoneally and the resulting CTL response was analysed after *in vitro* re-stimulation.

by competitive interference during *in vitro* re-stimulation and/or cytotoxicity assays. Therefore, the GP33-specific CTL response of lymphoma-bearing mice *in vivo* was evaluated by measuring strictly CTL-dependent LCMV-titre control. C57BL/6 mice primed with 10^5 EL4-GP33 cells injected intravenously eliminated the lymphoma (6 of 6 survivors) and controlled an LCMV infection below detectable levels (Table 2). In contrast, mice treated with 10^7 live EL4-GP33 cells could not control LCMV 14 days later and 6 of 6 died with tumours. Unresponsiveness was specific and not caused by generalized tumour growth. Mice primed against D^b-NP396 with vaccinia virus expressing LCMV-NP (Vacc-YN4; Table 2) 14 days before intravenous injection of 10^7 EL4-GP33 cells could control an LCMV infection. Control mice primed with LCMV 5 days before were all protected against both 10^7 EL4 lymphoma cells (day 0) and a subsequent LCMV infection (day 7). Thus, increased LCMV-GP-specific CTL frequencies prevented induction of specific unresponsiveness. These results extend earlier studies²¹ showing antigen-specific tolerance of CD4⁺ T cells to systemic tumours.

Cross processing of tumour antigens by host APCs through the class I MHC pathway (cross priming) has been postulated as a necessary and efficient mechanism to induce a protective CTL response^{12–16}. We tested cross priming with our panel of LCMV-NP- and LCMV-GP-expressing tumours or GP-expressing splenocytes. CB6 (BALB/c × C57BL/6) F₁ and B6D2 (C57BL/6 × B10.D2) F₁ mice were immunized with the various tumour cells used previously; in addition we used EL4-GP, transfected with the entire GP that was derived independently from the EL4-GP33 line.

MC-GP (H-2^b) cells induced a potent H-2^b-GP33- (Fig. 3A, b) and H-2^b-GP276-restricted (not shown) CTL response in H-2^b × H-2^d F₁ mice, but no H-2^d-GP283-specific CTLs (Fig. 3A, a). EL4-GP (H-2^b) induced H-2^b-GP33- but not H-2^d-restricted CTLs (Fig. 3A, e and f). In contrast, despite lower LCMV-GP expression when compared with MC-GP, D2-GP primed H-2^d-GP283-

restricted but not H-2^b-GP33-restricted CTLs (Fig. 3A, c versus d). Similarly, P815-NP (H-2^d) primed H-2^d NP118-specific but not H-2^b NP394-specific CTLs (Fig. 3A, g and h). Tumour cells (10^7) freeze thawed or heated at 42 °C, injected subcutaneously or intraperitoneally once or four times every second day, did not prime CTLs, but replicating or irradiated tumour cells induced CTLs directly and efficiently; therefore cross priming was undetectable. However, cross priming seems more efficient with antigen-expressing splenocytes than with tumour cells¹⁶. Therefore, LCMV-GP-expressing transgenic C57BL/6 (H-2^b, DEE) mice²² were bred with B10.D2 mice to obtain DEE^{b/d} F₁ and DEE^{d/d} F₂ offspring. Immunization of H-2^b × H-2^d F₁ mice with 2×10^6 DEE^{b/b} or DEE^{d/d} splenocytes induced CTLs restricted only to H-2^b-GP33 (Fig. 3B, a and b), or only H-2^d-GP283 (Fig. 1B, e and f), respectively. Control F₁ mice immunized with DEE^{b/d} had H-2^d- and H-2^b-restricted CTLs (Fig. 3B, c and d). Thus, GP-expressing splenocytes induced CTLs directly, but cross priming was undetectable.

Inefficient cross priming could be due to inefficient uptake by host APCs, or could reflect a relatively strict separation of MHC class I from class II pathways of antigen presentation or an unlikely selected 'avoidance of cross priming' by viral antigens. To test for conventional APC peptide presentation of exogenous cell debris through MHC class II, CB6 mice were immunized with 2×10^6 DEE^{d/d} splenocytes (or GP-expressing tumour cells; not shown). Twelve days later, these mice were challenged with the defined H-2^b T helper epitope P13 (GP60-80) coupled to 2,4-dinitrophenol (DNP)²³. DEE^{d/d}-primed, but not naïve, CB6 mice mounted DNP-specific immunoglobulin-γ (IgG) titres at day 7 (Fig. 3C, a). Therefore, LCMV-GP from DEE^{d/d} splenocytes was presented efficiently by MHC class II but not class I molecules of host APCs. Importantly, anti-CD4 treatment of mice revealed the necessity for CD4⁺ T-helper cells for direct CD8⁺ T-cell priming by all tumour cells (not shown) and DEE splenocytes (Fig. 3C, b). Therefore, antigen processing by host APCs seems necessary for

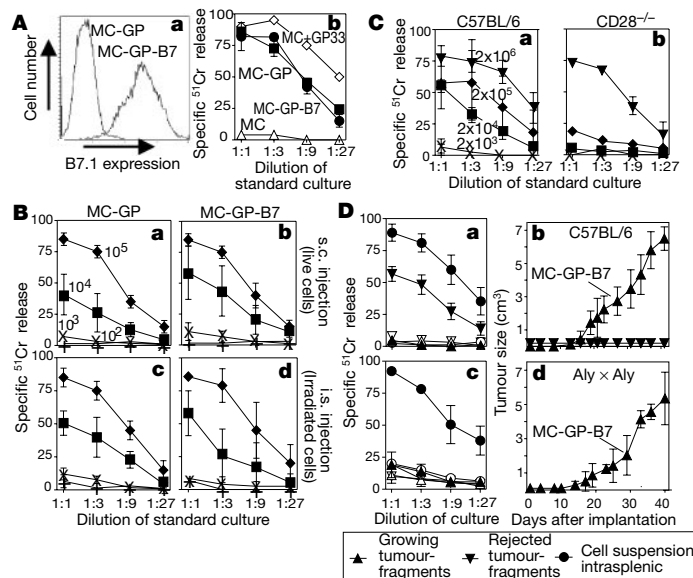


Figure 4 Role of B7 in the induction of an anti-tumour CTL response. **A**, MC-GP and MC-GP-B7 cells tested for B7.1 expression by FACS (**a**) and with or without peptide labelling (**b**) (GP33) as target cells in CTL assays. **B**, Immunogenicity of MC-GP versus MC-GP-B7 cells. **a–d**, 10^2 to 10^5 live (**a**, **b**) or irradiated (**c**, **d**) (80 Gy) MC-GP or MC-GP-B7 cells injected directly into the spleen in C57BL/6 mice; 8 days later secondary GP33-specific CTLs were tested. **C**, MC-GP priming in CD28^{-/-} mice. **a**, **b**, Titrated numbers of MC-GP cells injected intraperitoneally into C57BL/6 (**a**) or CD28^{-/-} (**b**) mice and tested for primed CTLs as in **A**. **D**, CTL induction and tumour growth in C57BL/6 and Aly × Aly

(C57BL/6 mice lacking all lymph nodes) recipients of tumour pieces. **a–d**, Secondary CTL activity (**a**, **c**) and tumour growth (**b**, **d**) presented as either growing (filled, normal triangle) or rejected (filled, inverted triangle) tumours. Secondary CTL responses in all mice at the end of the experiment (day 40) as in **A**. In addition, 10^5 MC-GP-B7 tumour cells were injected directly into the spleen of C57BL/6 (filled circles; **a**) and Aly × Aly mice (filled circles; **c**); 8 days later the secondary GP33-specific CTL activity was assessed as in **A**. Open symbols, unlabelled targets; filled symbols, GP33-labelled targets. Mean $\pm$ s.d. of 3–5 animals per group; 2–3 experiments with similar results.

induction of limiting T help, which is necessary for direct CTL priming by tumour cells. These findings confirm, extend and may explain both earlier experiments and some discrepancies on cross priming *in vivo*; they imply a limiting role for T-helper cells during priming *in vivo*¹⁸ but not for *in vitro* secondary CTL responses^{13,17}.

The two-signal hypothesis^{4–8} proposes that tumours without B7 or other co-stimulatory signals not only fail to induce CTLs but delete them. It predicts that strictly extra-lymphatic antigenic tumours expressing B7 should induce CTLs. We tested this hypothesis with MC-GP cells additionally transfected with B7.1, MC-GP-B7 (Fig. 4A, a) or with MC-GP cells without B7.1 in C57BL/6 mice possessing or lacking CD28 (ref. 24), the ligand for B7.1. MC-GP and MC-GP-B7 were comparably lysed by LCMV-specific CTLs (Fig. 4A, b). At least 10^4 live or irradiated MC-GP (Fig. 4B, a) or MC-GP-B7 (Fig. 4B, b) cells injected subcutaneously or into the spleen, respectively, were necessary to induce CTLs. To exclude small proliferation differences, CTL induction was compared after injection of irradiated MC-GP or MC-GP-B7 cells (80 Gy) subcutaneously or directly into the spleen. As with irradiated MC-GP cells, one dose of 10^7 irradiated MC-GP-B7 cells injected subcutaneously did not prime CTLs; however, $\geq 10^4$ irradiated MC-GP (Fig. 4B, c) or MC-GP-B7 cells (Fig. 4B, b) did prime CTLs when injected intrasplenically (see also Supplementary Information Fig. 2).

The overall role of B7–CD28 interactions in lymphoid organs was analysed 8 days after priming of CD28^{−/−} mice with different numbers of MC-GP cells. CD28^{−/−} mice generated CTLs; however, 100 times more MC-GP or MC-GP-B7 cells (not shown) injected intraperitoneally were necessary (Fig. 4C, a and b) compared with C57BL/6 mice. Taken together, B7 on tumour cells did not enhance induction of CTLs, whereas lack of CD28–B7 interactions in host lymphatic organs impaired CTL priming in a dose-dependent manner.

How then can other^{5–7} beneficial effects of B7 on tumour rejection be explained? MC-GP and MC-GP-B7 cell suspensions injected subcutaneously induced CTLs (Fig. 4B, a and b) and no

tumours grew. In contrast, small MC-GP-B7 tumour pieces transplanted subcutaneously grew; both MC-GP (not shown, but similar to Fig. 4D, b) and MC-GP-B7 (Fig. 4D, b) reached sizes of up to 5–7 cm³ after 30–40 days and did not prime CTLs (Fig. 4D, a). As shown above (Fig. 1B), mice that rejected tumours always exhibited primed CTLs independently of B7 expression on the tumour (Fig. 4D, a and b); again, all growing tumours expressed D^b-GP33 and B7 (not shown).

To avoid the migration of tumour cells to lymph nodes completely, cell suspensions or small tumour fragments were given to alymphoplastic (Aly × Aly) (C57BL/6) mice that lack lymph nodes but possess lymphatic vessels, spleen and normal CTLs²⁵. In contrast to controls, not only fragments but also cell suspensions of B7[−] or B7⁺ tumours injected subcutaneously grew in Aly × Aly mice (suspensions (2×10^6 cells): MC-GP 6 of 6 but also 10^4 cells formed tumours, MC-GP-B7 12 of 12; fragments: MC-GP 6 of 6, MC-GP-B7 8 of 8) and failed to prime splenic CTLs (Fig. 4D, c and d). Strictly peripheral MC-GP-B7 cells did not prime CTLs because they could not reach absent lymph nodes in Aly × Aly mice. Aly × Aly mice with growing tumours did generate CTLs when treated intrasplenically with 2×10^6 MC-GP-B7 cells (Fig. 4D, c), and could reject tumours when transfused with sufficient activated specific CTLs (Table 3). Thus CTLs were not rendered tolerant in tumour-bearing Aly × Aly mice; tumours were accessible and not generally suppressive. Even allogenic melanoma fragments (K1735, H-2^k) transfected or not with B7.1 (ref. 5) grew subcutaneously in Aly × Aly mice (10 of 10 K1735-B7 tumours grew) and no allospecific CTLs were induced; no ($n = 8$) tumours grew in C57BL/6 mice. Thus, an allogenic B7⁺ tumour strictly outside of lymphoid organs failed to prime CTLs.

To explain earlier documented benefits of B7 expression^{5–7} on immune rejection of tumours, we tested whether B7 expression on strictly peripheral tumours could enhance or maintain CTLs that had been already induced in lymphatic organs^{26,27}. MC-GP and MC-GP-B7 pieces were transplanted subcutaneously into the contralateral flank of the same T- and B-cell-deficient Rag-1^{−/−} mice—the tumours grew to a comparable size (5–7 mm) by day 10 after transplantation. At this time point, 5×10^7 GP33-specific acute day 8 LCMV immune spleen cells were transfused into tumour-bearing Rag-1^{−/−} mice. Only 1 out of 7 MC-GP but 6 of 7 MC-GP-B7 tumours were controlled in the same recipient mice (Table 3).

Immunohistology showed similar CD4⁺ T-cell, B-cell or CD11⁺ APC infiltration in B7⁺ versus B7[−] tumours (not shown). In contrast, CD8⁺ T-cell infiltration in MC-GP tumours was more extensive in connective tissue around blood vessels and around the border of solid tumour tissue, but they infiltrated MC-GP-B7 tumours efficiently (Fig. 5).

Our findings suggest that strictly extra-lymphatic peripheral tumours are ignored by CTLs and do not delete them. Perhaps rare tumours that reach lymphatic organs early are rejected before they become clinically evident. Once a tumour has reached a certain size and effector CTLs get induced late, the outcome reflects a race of

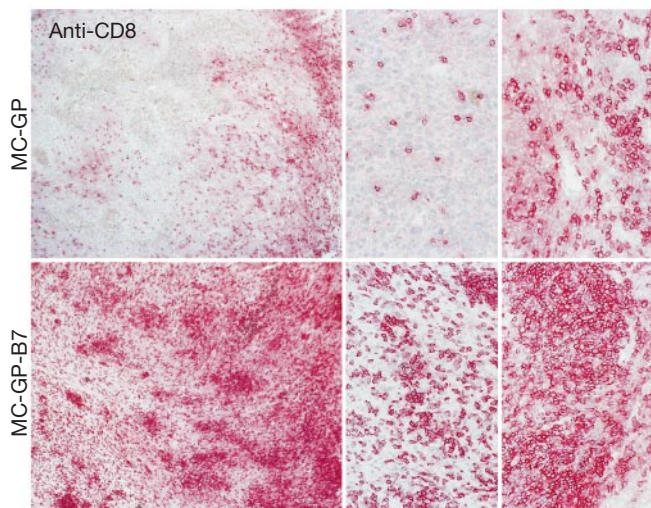


Figure 5 Immunohistochemistry of CTL-infiltrated MC-GP and MC-GP-B7 tumours. Tumour fragments of MC-GP and MC-GP-B7 tumours implanted subcutaneously each on one flank of Rag-1^{−/−} mice. Ten days later 5×10^7 day-8 LCMV effector cells (splenocytes of C57BL/6 mice infected 8 days previously with 200 PFU LCMV) were adoptively transferred into tumour-bearing recipients. Ten days later, tumours were removed and immunohistology was performed with anti-CD8. Magnification: $\times 50$, left panels middle panels (middle of the tumour); $\times 200$, middle panels (middle of the tumour) and right panels (at the tumour border). One out of five comparable histological sections for each tumour is shown.

Table 3 Control of B7⁺ and B7[−] tumours by transferred specific CTLs

Recipient	Tumour transplant*	LCMV CTLs†	Controlled/tested‡
Aly × Aly	MC-GP	—	0/6
Aly × Aly	MC-GP	$3 \times 2 \times 10^7$	7/8
Rag-1 ^{−/−}	MC-GP	—	0/6
Rag-1 ^{−/−}	MC-GP-B7	—	0/6
Rag-1 ^{−/−}	MC-GP	5×10^7	1/7
Rag-1 ^{−/−}	MC-GP-B7	5×10^7	6/7

*Tumours were transplanted subcutaneously as solid fragments of MC-GP and MC-GP-B7 into contralateral flanks of the same recipient mouse 14 days before adoptive transfer.

† Splenocytes of C57BL/6 mice immunized with 200 PFU LCMV (8 days previously) and transferred (i.v.) to mice bearing tumours ≈ 0.5 cm. Tumour control in Aly × Aly mice after repetitive (3 times in 3-day intervals) transfer of 2×10^7 effector cells.

‡ Tumour control was defined as complete rejection or stable tumour size for the next 21 days.

growing tumour masses and limited effector cells versus tumour mutant selection. One extreme form of tumour selection in the host is probably illustrated here by lymphohaemopoietic tumours where exhaustive induction and deletion of CTLs correlated with the tumour success.

Our results seem to contradict the fact that late lymph-node metastases usually signal failure of immunosurveillance^{1,2}. We show that very early trapping of tumour cells in secondary lymphoid organs is beneficial as it induces CTLs to reject tiny tumours early, at a time when selection of immune evasion mutants is improbable.

In contrast to earlier studies^{12–16} our analyses show direct, specific CTL induction by signal 2 negative tumour cells without demonstrable cross presentation through MHC class I when analysed for peptide specificity *ex vivo* or protection *in vivo*. Notably, direct CTL induction depended on immune CD4⁺ T-helper-cell activity^{18,28}. As fibroblasts and epithelial cells are MHC class II-negative, antigen presentation to CD4⁺ cells depends on and is performed conventionally by host APCs. Perhaps some results that have not defined the cross-priming peptide specificity may indicate T-helper cell involvement during priming *in vivo*^{17,28}. Cross priming has been documented *in vitro*¹⁴ and *in vivo* with ovalbumin and viral antigens^{12,13,17} but its role in CTL-mediated protection is unclear. Our study suggests that such a process seems inefficient unless special tricks are used^{12–15,17}. Only additional studies will show whether antigen studied under special conditions or with normal or very high CTL frequencies reveal generalizable versus rare mechanisms of cross priming of CTL induction or tolerization *in vivo*^{12–15,17}.

CTL induction in secondary lymphoid tissues, while enhanced by CD28, was independent of B7.1 on tumour cells. Thus CD28–B7 co-stimulation acts in bystander fashion in lymphatic organs but improved or maintained CTL effectors in B7⁺ peripheral solid tumours—other results are compatible with our findings^{26,27}. Use of transgenic T cells may yield uninterpretable results because their high frequency enhances bystander induction or activation in secondary lymphoid organs, faking peripheral B7-driven induction²⁹. Our results also document that neither B7⁺ nor B7[−] strictly extra-lymphatic tumours can induce or delete T cells, negating predictions of the two-signal hypothesis. Therefore, T-cell deletion by tumours is probably not through negative signals but more likely represents 'over-induction', suggested for CD4⁺ T cells²¹ or here for CTLs. An involvement of B7 on natural killer cells³⁰ was not analysed here, but B7-independent growth kinetics and the lack of natural killer control of MC-GP tumours¹⁰ renders this possibility unlikely.

This study suggests a number of apparently simple possibilities to improve the overall unfavourable balance between peripheral solid tumours and the host. They all depend on tumour loads being as small as possible at an early time of T-cell induction in lymphoid organs to avoid immune escape, and also on the maintenance of great numbers of activated T cells. □

Methods

Mice and *in vivo* experiments

C57BL/6, lymph-node-deficient Aly×Aly²⁵ and DEE mice²² (LCMV-GP transgenic) on a pure C57BL/6 background were bred locally. We obtained CD28^{−/−} mice²⁴ from the Basel Institute for Immunology. Tumour size was assessed twice a week and the animals were killed when the tumour reached 6 cm³. Tumour volume was calculated by the formula $V = \pi \times abc/6$, where a , b and c are the orthogonal diameters.

Tumour cell lines, cell analysis, histology and CTL assays

The MC57-GP^{9,10}, LL-GP33 (ref. 19) and B16-GP33 (ref. 20) cell lines were described earlier. A part of the LCMV-GP (residues 1–60) as well as the entire LCMV-GP and LCMV-NP were subcloned from the pLCMV-GP and pLCMV-NP plasmids, respectively, into a CMV-driven eukaryotic expression vector containing the geneticin resistance gene. Co-transfection⁹ of MC-GP with a plasmid encoding the murine B7.1 (CMV promoter) and a hygromycin-resistance gene (pX343) (Hygromycin B 150 µg ml^{−1}; Fluka Chemie AG) by calcium phosphate co-precipitation. The K1735 B7.1⁺ or B7.1[−] cell lines were a gift of J. P. Allison⁵.

All FACS antibodies were from PharMingen. Histochemistry was performed as described²⁵. Histological collagen staining was done with sirius red. *In vitro* re-stimulation of LCMV GP- or NP-specific or of allospecific spleen cells has been reported previously²⁵.

Received 20 November 2000; accepted 3 May 2001.

- Old, L. J. Tumour immunology: the first century. *Curr. Opin. Immunol.* **4**, 603–607 (1992).
- Boon, T., Cerottini, J. C., van-den-Eynde, B., van-der-Bruggen, P. & Van-Pel, A. Tumour antigens recognised by T lymphocytes. *Annu. Rev. Immunol.* **12**, 337–365 (1994).
- Braun, S. *et al.* Cytokeratin-positive cells in the bone marrow and survival of patients with stage I, II or III breast cancer. *N. Engl. J. Med.* **342**, 525–533 (2000).
- Schwartz, R. H. Costimulation of T lymphocytes: the role of CD28, CTLA-4, and B7/BB1 in interleukin-2 production and immunotherapy. *Cell* **71**, 1065–1068 (1992).
- Townsend, S. E. & Allison, J. P. Tumour rejection after direct costimulation of CD8⁺ T cells by B7-transfected melanoma cells. *Science* **259**, 368–370 (1993).
- Chen, L. *et al.* Costimulation of antitumour immunity by the B7 counterreceptor for the T lymphocyte molecules CD28 and CTLA-4. *Cell* **71**, 1093–1102 (1992).
- Hellstrom, K. E., Hellstrom, I. & Chen, L. Can co-stimulated tumour immunity be therapeutically efficacious? *Immunol. Rev.* **145**, 123–145 (1995).
- Matzinger, P. Tolerance, danger, and the extended family. *Annu. Rev. Immunol.* **12**, 991–1045 (1994).
- Kündig, T. M. *et al.* Fibroblasts as efficient antigen-presenting cells in lymphoid organs. *Science* **268**, 1343–1347 (1995).
- Ochsenbein, A. F. *et al.* Immune surveillance against a peripheral solid tumour fails because of immunological ignorance. *Proc. Natl Acad. Sci. USA* **96**, 2233–2238 (1999).
- Hahn, H. & Kaufmann, S. H. E. Role of cell mediated immunity in bacterial infections. *Rev. Infect. Dis.* **3**, 1221–1250 (1981).
- Carbone, F. R., Kurts, C., Bennet, S. R. M., Miller, J. F. A. P. & Heath, W. R. Cross-presentation: a general mechanism for CTL immunity and tolerance. *Immunol. Today* **19**, 368–373 (1998).
- Huang, A. Y. *et al.* Role of bone marrow-derived cells in presenting MHC class I-restricted tumour antigens. *Science* **264**, 961–965 (1994).
- Yewdell, J. W., Norbury, C. C. & Bennink, J. R. Mechanisms of exogenous antigen presentation by MHC class I molecules *in vitro* and *in vivo*: implications for generating CD8⁺ T cell responses to infectious agents, tumours, transplants, and vaccines. *Adv. Immunol.* **73**, 1–77 (1999).
- Jeannin, P. OmpA targets dendritic cells, induces their maturation and delivers antigen into the MHC class I presentation pathway. *Nature Immunol.* **1**, 502–509 (2000).
- Bevan, M. J. Cross-priming for a secondary cytotoxic response to minor H antigens with H-2 congenic cells which do not cross-react in the cytotoxic assay. *J. Exp. Med.* **143**, 1283–1288 (1976).
- Toes, R. E. M. *et al.* Protective antitumour immunity induced by immunization with completely allogenic tumour cells. *Cancer Res.* **56**, 3782–3787 (1996).
- Lu, Z. *et al.* CD40-independent pathways of T cell help for priming of CD8⁺ cytotoxic T lymphocytes. *J. Exp. Med.* **191**, 541–550 (2000).
- Hermans, I. F., Daish, A., Yang, J., Ritchie, D. S. & Ronchese, F. Antigen expressed on tumour cells fails to elicit an immune response, even in the presence of increased numbers of tumour-specific cytotoxic T lymphocyte precursors. *Cancer Res.* **58**, 3909–3917 (1998).
- Prevost-Blondel, A. *et al.* Tumour-infiltrating lymphocytes exhibiting high *ex vivo* cytolytic activity fail to prevent murine melanoma tumour growth *in vivo*. *J. Immunol.* **161**, 2187–2194 (1998).
- Staveley-O'Carroll, K. *et al.* Induction of antigen-specific T cell anergy: an early event in the course of tumour progression. *Proc. Natl Acad. Sci. USA* **95**, 1178–1183 (1998).
- Oehen, S. U. *et al.* Escape of thymocytes and mature T cells from clonal deletion due to limiting tolerogen expression levels. *Cell Immunol.* **158**, 342–352 (1994).
- Oxenius, A., Bachmann, M. F., Zinkernagel, R. M. & Hengartner, H. Virus-specific MHC class II-restricted TCR-transgenic mice: effects on humoral and cellular immune response after viral infection. *Eur. J. Immunol.* **28**, 390–400 (1998).
- Shahinian, A. *et al.* Differential T cell costimulatory requirements in CD28-deficient mice. *Science* **261**, 609–612 (1993).
- Karrer, U. *et al.* On the key role of secondary lymphoid organs in antiviral immune responses studied in alymphoplastic (aly/aly) and spleenless (Hox11^{−/−}) mutant mice. *J. Exp. Med.* **185**, 2157–2170 (1997).
- Ramarathinam, L., Castle, M., Wu, Y. & Liu, Y. T cell costimulation by B7/BB1 induces CD8 T cell-dependent tumour rejection: an important role of B7/BB1 in the induction, recruitment, and effector function of antitumour T cells. *J. Exp. Med.* **179**, 1205–1214 (1994).
- Maric, M., Zheng, P., Sarma, S., Guo, Y. & Liu, Y. Maturation of cytotoxic T lymphocytes against a B7-transfected nonmetastatic tumour: a critical role for costimulation by B7 on both tumour and host antigen-presenting cells. *Cancer Res.* **58**, 3376–3384 (1998).
- Ossendrop, F., Mengede, E., Camps, M., Filius, R. & Melief, C. J. Specific T helper requirement for optimal induction of cytotoxic T lymphocytes against major histocompatibility complex class II negative tumours. *J. Exp. Med.* **187**, 693–702 (1998).
- Harlan, D. M. *et al.* Mice expressing both B7-1 and viral glycoprotein on pancreatic beta cells along with glycoprotein-specific transgenic T cells develop diabetes due to a breakdown of T-lymphocyte unresponsiveness. *Proc. Natl Acad. Sci. USA* **91**, 3137–3141 (1994).
- Wu, T. C., Huang, A. Y. C., Jaffee, E. M., Levitsky, H. I. & Pardoll, D. M. A reassessment of the role of B7-1 expression in tumour rejection. *J. Exp. Med.* **182**, 1415–1421 (1995).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank K. McCoy for critical review of the manuscript and N. Wey for photographs. This work was funded by grants from the Swiss National Science Foundation (to R.M.Z.) and the Kanton Zurich.

Correspondence and requests for material should be addressed to R.M.Z. (e-mail: rolf.zinkernagel@pty.usz.ch) or A.F.O. (e-mail: adrian.ochsenbein@insel.ch).

Spatio-temporal images of growth-factor-induced activation of Ras and Rap1

Naoki Mochizuki*, Shigeko Yamashita†, Kazuo Kurokawa‡, Yusuke Ohba‡, Takeharu Nagai§, Atsushi Miyawaki§ & Michiyuki Matsuda‡

* Department of Structural Analysis, National Cardiovascular Center Research Institute, 5-7-1 Fujishirodai, Suita-shi, Osaka 565-8565, Japan

† Department of Pathology, Research Institute, International Medical Center of Japan, 1-21-1 Toyama, Shinjuku-ku, Tokyo 162-8655, Japan

‡ Department of Tumor Virology, Institute for Microbial Diseases, Osaka University, 3-1 Yamadaoka, Suita-shi, Osaka 565-0871, Japan

§ Laboratory for Cell Function and Dynamics, Brain Science Institute, RIKEN, 2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan

G proteins of the Ras family function as molecular switches in many signalling cascades^{1–3}; however, little is known about where they become activated in living cells. Here we use FRET (fluorescent resonance energy transfer)-based sensors to report on the spatio-temporal images of growth-factor-induced activation of Ras and Rap1. Epidermal growth factor activated Ras at the peripheral plasma membrane and Rap1 at the intracellular perinuclear region of COS-1 cells. In PC12 cells, nerve growth factor-induced activation of Ras was initiated at the plasma membrane and transmitted to the whole cell body. After three hours, high Ras activity was observed at the extending neurites. By using the FRAP (fluorescence recovery after photobleaching) technique, we found that Ras at the neurites turned over rapidly; therefore, the

sustained Ras activity at neurites was due to high GTP/GDP exchange rate and/or low GTPase activity, but not to the retention of the active Ras. These observations may resolve long-standing questions as to how Ras and Rap1 induce different cellular responses⁴ and how the signals for differentiation and survival are distinguished by neuronal cells⁵.

G proteins of the Ras family cycle between GDP-bound inactive and GTP-bound active forms. This cycle is regulated by guanine nucleotide exchange factor (GEF), the activator, and GTPase activating protein (GAP), the inactivator. Many external and internal signals converge on G proteins of the Ras family through several GEFs and GAPs; however, if so many signals can activate G proteins of the Ras family, it remains unknown how they transmit a specific signal^{4,6}.

After making a calcium indicator based on FRET technology⁷, we designed a protein that consisted of H-Ras, the Ras-binding domain of Raf (RafRBD) and a pair of an enhanced yellow-emitting mutant of green fluorescent protein (YFP) and an enhanced cyan-emitting mutant of green fluorescent protein (CFP), so that intramolecular binding of GTP-Ras to Raf RBD brings CFP close to YFP and increases FRET between CFP and YFP (Fig. 1a). This protein was named a Ras and interacting protein chimaeric unit (Raichu). By replacing Ras with Rap1, we also made a probe for Rap1. Raichu-Ras and Raichu-Rap1 were co-expressed in 293T cells with either GEFs or GAPs, and their emission profiles were examined with an excitation wavelength of 433 nm (Fig. 1b). FRET, which was typically observed as an emission peak of 527 nm, was more prominent in the presence of GEFs than in the presence of GAPs. After treatment with proteinase K, which cleaved Raichu-Ras and Raichu-Rap1 between YFP and CFP, the emission peak at 527 nm disappeared, proving that this peak was caused by FRET⁸.

To correlate GTP loading with FRET efficiency, we labelled cells expressing Raichu-Ras, Flag-tagged H-Ras and a varying quantity

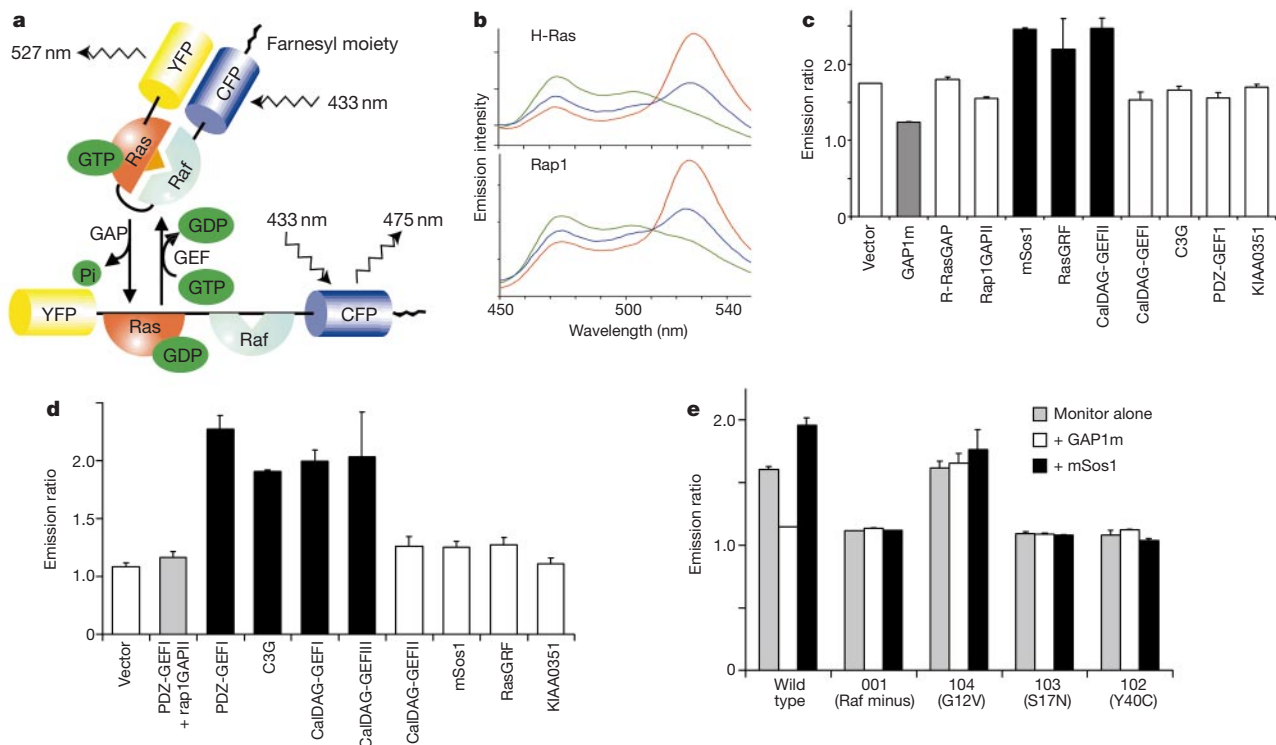


Figure 1 Properties of Raichu-Ras and Raichu-Rap1. **a**, Schematic representation of Raichu-Ras bound to GDP or GTP. **b**, Emission spectra of Raichu-Ras and Raichu-Rap1 (excited at 433 nm) co-expressed with GEF (red) or GAP (blue) in 293T cells. To confirm FRET, one sample was treated with proteinase K (green), which cleaved Raichu between

CFP and YFP. **c**, **d**, Emission ratio of Raichu-Ras or Raichu-Rap1 co-expressed with GAPs and GEFs in 293T cells. GAPs and GEFs for Ras (**c**) or Rap1 (**d**) are indicated in grey and black columns, respectively. **e**, Emission ratio of Raichu-Ras mutants expressed with or without GAP1m or mSos1 in 293T cells. Bars are s.d. ($n = 3$).

of either mSos1 or GAP1m with $^{32}\text{P}_i$ and determined the amount of GTP bound to Raichu-Ras and Flag-tagged H-Ras by thin layer chromatography (see Supplementary Information Fig. 1). FRET efficiency, which is shown as an emission ratio of 527/475 nm, correlated well with the amount of GTP bound to Raichu-Ras. The percentage of GTP-bound Raichu-Ras was always higher than that of GTP-bound Flag-H-Ras obtained from the same cells. This is probably because Raf RBD present on Raichu-Ras competitively inhibits GAP⁹. Despite this, the linear correlation of GTP-loading and FRET efficiency and the large dynamic range of the emission ratio from 1.2 to 2.4 show that Raichu-Ras can be used for monitoring Ras activation *in vivo*. Similar results were obtained for Raichu-Rap1.

We next tested the specificity of Raichu-Ras on a panel of GEFs and GAPs of G proteins of the Ras family (Fig. 1c). Neither R-RasGAP, which is related closely to GAP1m but specific to R-Ras, nor rap1GAP1, a GAP for Rap1, stimulated the GTPase activity of Raichu-Ras. RasGRF and CalDAG-GEFII, GEFs for Ras and R-Ras, increased FRET, as did mSos1. In contrast, GEFs for Rap1, R-Ras or Ral including CalDAG-GEFI, C3G, PDZ-GEFI and KIAA0351 (RalGPS/RalGEF2) did not show any effect on the FRET efficiency of Raichu-Ras. Thus, Raichu-Ras retained specificity for both GEFs and GAPs of the classical Ras subfamily. Again, similar results were obtained for Raichu-Rap1 (Fig. 1d): the emission ratio of Raichu-Rap1 was increased in the presence of GEFs for Rap1 including PDZ-GEFI, C3G, CalDAG-GEFI and CalDAG-GEFIII, and was decreased in the presence of GAPs for G proteins of the Rap family including rap1GAP1 and Spa1 (see Supplementary Information Fig. 2).

To confirm that the increased emission ratio of Raichu-Ras was caused by the intramolecular binding of Ras to Raf RBD, we constructed a series of mutants: Raichu 001 lacked Raf RBD; Raichu 103 contained an S17N mutation in Ras, which reduced the affinity to guanine nucleotides¹⁰; and Raichu 102 contained a Y40C mutation in which the interaction of H-Ras with Raf was abolished¹¹. In these mutants, the FRET efficiency was not enhanced

in the presence of mSos1 (Fig. 1e). Raichu 104, which lacked GTPase activity because of a G12V mutation in Ras², had an increased FRET efficiency irrespective of the presence of GAP1m or mSos1. Thus, the effects of mutations shown either to inhibit or to activate H-Ras were reproduced upon mutation of Raichu-Ras.

Because the effector domain of the GTP-loaded Raichu is occupied by Raf RBD, Raichu will not perturb the cellular signalling cascades by binding to the cellular target proteins. In fact, we confirmed that expression of Raichu-Ras and Raichu-Rap1 neither enhanced nor attenuated the epidermal growth factor (EGF)-induced ERK/MAPK activation (see Supplementary Information Fig. 3).

With these probes, we monitored the activation of Ras and Rap1 in living COS-1 cells by a dual-emission ratio fluorescent microscopy. CFP and YFP images were taken every 20 s and used to create FRET images (see Supplementary Information movie 1). In the following figures, the FRET images are presented in intensity-modulated display (IMD) mode, which associates colour hue with emission ratio value and the intensity of each hue with the source image brightness. We found that EGF-induced Ras activation was marked at the periphery of the cells, where membrane ruffling was prominent (Fig. 2a and Supplementary Information movie 2). In contrast, EGF activated Rap1 mostly in the centre of the cells (Fig. 2a and Supplementary Information movie 2). Use of a confocal microscope showed more clearly that, upon EGF stimulation, Ras activity increased mostly at the peripheral plasma membrane (Fig. 2b). However, EGF activated Rap1 at the internal perinuclear region of the cells, but not at the plasma membrane. EGF did not stimulate a Raichu-Rap1 mutant that lacked a farnesyl moiety,

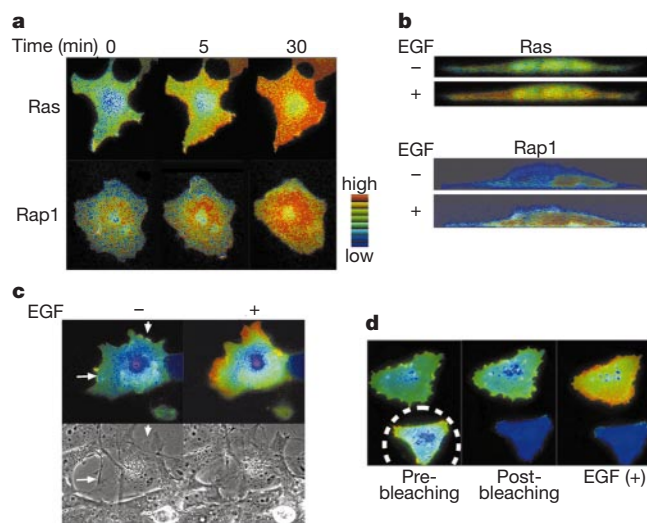


Figure 2 EGF activation of Ras and Rap1. **a**, IMD images of COS-1 cells expressing Raichu-Ras or Raichu-Rap1 and stimulated by EGF. **b**, Tangential IMD images obtained with a confocal microscope. COS-1 cell expressing Raichu-Ras or Raichu-Rap1 were stimulated by EGF for 5 min. **c**, Phase contrast and IMD images of a Raichu-Ras-expressing COS-1 cell, which had grown to subconfluency and was stimulated by EGF for 5 min. Arrows indicate the free edge of the cell, to which Ras activation was restricted. **d**, Cell indicated by a dotted line (left) was photobleached at an excitation wavelength of 510 nm for 2 min (centre) and stimulated by EGF for 5 min (right).

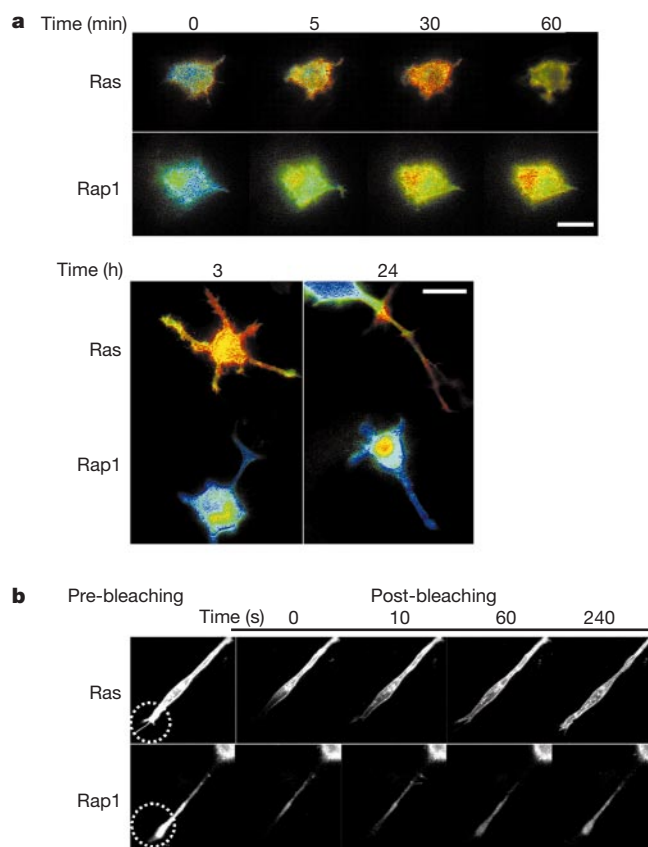


Figure 3 Activation of Ras and Rap1 in NGF-stimulated PC12 cells. **a**, IMD images of PC12 cells expressing Raichu-Ras or Raichu-Rap1 and stimulated by NGF. Scale bars, 10 μm . **b**, Neurites of differentiated PC12 cells expressing YFP-Ras or YFP-Rap1 after they were photobleached.

indicating that Rap1 was activated at the intracellular membrane compartments. When we pretreated cells with monodansylcadaverine, an inhibitor for clathrin-mediated endocytosis, EGF activated Ras, but not Rap1. Therefore, only the internalized EGF receptor may activate C3G, a Rap1 GEF activated by tyrosine phosphorylation¹².

When subconfluent COS-1 cells were stimulated with EGF, Ras was activated mostly at the free edge of the cells, but less at the edge in contact with the neighbouring cells, where membrane ruffling was suppressed (Fig. 2c). It is proposed that GAP activity is increased in confluent cells¹³. Our observation indicates that local activation of GAP at the sites of cell-to-cell contact may attenuate Ras activation.

The EGF-induced increase in FRET was not observed in Raichu-102 (Y40C mutant)-expressing cells, in cells co-expressing Raichu-Ras with a dominant negative mutant of H-Ras (S17N) (see Supplementary Information movie 3) or in cells pretreated with AG1478, a tyrosine kinase inhibitor. To confirm that the increase in the emission ratio of YFP to CFP was due to FRET, we photobleached one of a pair of Raichu-Ras-expressing cells on YFP before EGF stimulation⁸. As expected, only the non-photobleached cells had an increased emission ratio upon EGF stimulation (Fig. 2d).

We next examined the activation of Ras and Rap1 in the rat pheochromocytoma cell line PC12. Upon nerve growth factor (NGF) stimulation, Ras was activated from the periphery of the PC12 cells (Fig. 3a). Within 3 h, Ras activation diminished in the cell body and became apparent in the extending neurites. After 24 h, when PC12 cells differentiated into neuronal cells, activation of Ras persisted only in the extended neurites. Again, in contrast to Ras, Rap1 was activated at the intracellular region of NGF-stimulated PC12 cells, but never at the neurites. It has been proposed that differentiation of the PC12 cells is prompted by the NGF receptor (TrkA) at the endosomes and that survival responses are initiated by TrkA at the cell surface¹⁴. This is consistent with our observation that Ras is activated transiently at the cell body and persistently at the extended neurites. The high Ras activity at the neurites may be maintained by the NGF-TrkA complex, which is transported from the axon to the cells and is required for ERK-dependent phosphorylation of CREB cyclic AMP response element-binding protein¹⁵.

To understand the mechanism underlying the intracellular gradient of the activities of Ras and Rap1, the mobility of Ras and Rap1 was examined by FRAP¹⁶. We photobleached neurites of the PC12 cells expressing YFP-Ras or YFP-Rap1 and monitored the recovery of fluorescence intensity (Fig. 3b). In both cells, the fluorescence intensity recovered within 240 s. This observation indicates that the sustained high activity of Ras and low activity of Rap1 were not caused by the retention of active Ras or inactive Rap1 at the neurites. It is likely that local balance between the activities of GEFs and GAPs determines the local activities of Ras and Rap1.

Rap1 antagonizes Ras-dependent cellular transformation and ERK activation in some cell types^{17,18}, probably by sequestering the Ras-Raf complex¹⁹. We have found that, upon growth factor activation, the localizations of the active Ras and active Rap1 are exclusive of each other. Moreover, expression of active Rap1, Rap1V12, did not inhibit EGF-induced Ras activation (see Supplementary Information movie 3). Therefore, it is unlikely that activated Rap1 inhibits the Ras-dependent ERK activation by trapping the Ras-Raf complex. However, Rap1 may attenuate Ras-dependent activation of ERK by retaining Raf in the perinuclear region and reducing the quantity of Raf at the plasma membrane. This view is supported by the observation that overexpression of rap1GAP enhanced growth-factor-induced ERK activation²⁰.

Previous biochemical analyses rarely detected more than 40% activation of Ras and Rap1 upon growth factor stimulation²¹. It has now become clear that growth-factor stimulation strongly activates Ras and Rap1 in a very restricted area within cells, and that a large

population of Ras or Rap1 remains inactive, causing an apparent low-level response in biochemical assays. Thus, the probes reported here enable highly sensitive monitoring of the activities of Ras and Rap1 with spatio-temporal information. This property also promises the widespread application of this probe *in vivo*. □

Methods

For detailed methods, see Supplementary Information.

Plasmids

pRaichu-Ras was derived from the pCAGGS eukaryotic expression vector²² and encoded a chimeric protein, Raichu-Ras, which consisted of EYFP-V68L/Q69K²³, H-Ras, Raf RBD and ECFP from the amino terminus. In pRaichu-Rap1, H-Ras was replaced with Rap1A. The nucleotide sequences of the coding regions of pRaichu-Ras and pRaichu-Rap1 were deposited in the DDBJ/EMBL/GenBank data (accession nos AB046925 and AB051846). Plasmids used in this study have been described previously²⁴.

Cells and Imaging

COS-1 and PC12 cells were plated on a collagen-coated 35-mm-diameter glass-base dish (Asahi Techno Glass), transfected with plasmids, starved of serum for 6 h, and stimulated with 5 ng ml⁻¹ EGF or 25 ng ml⁻¹ NGF. In some experiments, cells were pretreated with 10 nM AG1478 (Calbiochem) for 30 min or 20 μM monodansylcadaverine (Sigma) for 10 min. Dual-emission ratio imaging and FRAP experiments were performed as described elsewhere^{7,25}.

Received 7 December 2000; accepted 26 March 2001.

- Barbacid, M. *ras* genes. *Annu. Rev. Biochem.* **56**, 779–827 (1987).
- Lowy, D. R. & Willumsen, B. M. Function and regulation of *ras*. *Annu. Rev. Biochem.* **62**, 851–891 (1993).
- Schlessinger, J. & Bar-Sagi, D. Activation of *ras* and other signaling pathways by receptor tyrosine kinases. *Cold Spring Harb. Symp. Quant. Biol.* **59**, 173–179 (1994).
- Bos, J. L. All in the family? New insights and questions regarding interconnectivity of Ras, Rap1 and Ral. *EMBO J.* **17**, 6776–6782 (1998).
- Marshall, C. J. Specificity of receptor tyrosine kinase signaling: transient versus sustained extracellular signal-regulated kinase activation. *Cell* **80**, 179–185 (1995).
- Vojtek, A. B. & Der, C. J. Increasing complexity of the Ras signaling pathway. *J. Biol. Chem.* **273**, 19925–19928 (1998).
- Miyawaki, A. *et al.* Fluorescent indicators for Ca²⁺ based on green fluorescent proteins and calmodulin. *Nature* **388**, 882–887 (1997).
- Miyawaki, A. & Tsien, R. Y. Monitoring protein conformations and interactions by fluorescence resonance energy transfer. *Methods Enzymol.* **327**, 472–500 (2000).
- Scheffler, J. E. *et al.* Characterization of a 78-residue fragment of c-Raf-1 that comprises a minimal binding domain for the interaction with Ras-GTP. *J. Biol. Chem.* **269**, 22340–22346 (1994).
- Feig, L. A. & Cooper, G. M. Inhibition of NIH3T3 cell proliferation by a mutant *ras* protein with preferential affinity for GDP. *Mol. Cell. Biol.* **8**, 3235–3243 (1988).
- Rodriguez-Viciana, P. *et al.* Role of phosphoinositide 3-OH kinase in cell transformation and control of the actin cytoskeleton by Ras. *Cell* **89**, 457–467 (1997).
- Ichiba, T. *et al.* Activation of C3G guanine nucleotide exchange factor for Rap1 by phosphorylation of tyrosine 504. *J. Biol. Chem.* **274**, 14376–14381 (1999).
- Hoshino, M., Kawakita, M. & Hattori, S. Characterization of a factor that stimulates hydrolysis of GTP bound to *ras* gene product p21 (GTPase-activating protein) and correlation of its activity to cell density. *Mol. Cell. Biol.* **8**, 4169–4173 (1988).
- Zhang, Y., Moheban, D. B., Conway, B. R., Bhattacharyya, A., & Segal, R. A. Cell surface Trk receptors mediate NGF-induced survival while internalized receptors regulate NGF-induced differentiation. *J. Neurosci.* **20**, 5671–5678 (2000).
- Riccio, A., Pierchala, B. A., Ciarallo, C. L. & Ginty, D. D. An NGF-TrkA-mediated retrograde signal to transcription factor CREB in sympathetic neurons. *Science* **277**, 1097–1100 (1997).
- Xing, J., Kornhauser, J. M., Xia, Z., Thiele, E. A. & Greenberg, M. E. Nerve growth factor activates extracellular signal-regulated kinase and p38 mitogen-activated protein kinase pathways to stimulate CREB serine 133 phosphorylation. *Mol. Cell Biol.* **18**, 1946–1955 (1998).
- Kitayama, H., Sugimoto, Y., Matsuzaki, T., Ikawa, Y. & Noda, M. A *ras*-related gene with transformation suppressor activity. *Cell* **56**, 77–84 (1989).
- Cook, S. J., Rubinfeld, B., Albert, I. & McCormick, F. RapV12 antagonizes Ras-dependent activation of ERK1 and ERK2 by LPA and EGF in Rat-1 fibroblasts. *EMBO J.* **12**, 3475–3485 (1993).
- Hu, C. D. *et al.* Coassociation of Rap1A and Ha-Ras with Raf-1 N-terminal region interferes with Ras-dependent activation of Raf-1. *J. Biol. Chem.* **272**, 11702–11705 (1997).
- Mochizuki, N. *et al.* Activation of ERK/MAPK pathway by an isoform of rap1GAP associated with G_{ai}. *Nature* **400**, 891–894 (1999).
- Satoh, T., Nakafuku, M. & Kaziro, Y. Function of Ras as a molecular switch in signal transduction. *J. Biol. Chem.* **267**, 24149–24152 (1992).
- Niwa, H., Yamamura, K. & Miyazaki, J. Efficient selection for high-expression transfectants with a novel eukaryotic vector. *Gene* **108**, 193–200 (1991).
- Miyawaki, A., Griesbeck, O., Heim, R. & Tsien, R. Y. Dynamic and quantitative Ca²⁺ measurements using improved green fluorescent proteins. *Proc. Natl Acad. Sci. USA* **96**, 2135–2140 (1999).
- Ohba, Y. *et al.* Rap2 as a slowly responding molecular switch in the Rap1 signaling cascade. *Mol. Cell Biol.* **20**, 6074–6083 (2000).
- Lippincott-Schwartz, J. *et al.* Monitoring the dynamics and mobility of membrane proteins tagged with green fluorescent protein. *Methods Cell Biol.* **58**, 261–281 (1999).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank D. Bowtell, K. Kaibuchi, J. Miyazaki, N. Minato and N. Nomura for plasmids; H. Mizuno, F. Ohba, N. Otsuka, K. Kimura and K. Okuda for technical assistance; and S. Hattori, M. Dutta and A. Dutta for the critical reading of the manuscript. Supported by grants from the Ministry of Health and Welfare, from the Ministry of Education, Science, Sports and Culture, Mitsui Life Social Welfare Foundation, the Princess Takamatsu Cancer Research Fund and from the Japan Health Science Foundation.

Correspondence and requests for materials should be addressed to M.M. (e-mail: matsudam@biken.osaka-u.ac.jp).

Cyclin-dependent kinases prevent DNA re-replication through multiple mechanisms

Van Q. Nguyen*, Carl Co* & Joachim J. Li†

* Department of Biochemistry and Biophysics; and † Department of Microbiology and Immunology, University of California, San Francisco, California 94143-0414, USA

The stable propagation of genetic information requires that the entire genome of an organism be faithfully replicated once and only once each cell cycle. In eukaryotes, this replication is initiated at hundreds to thousands of replication origins distributed over

the genome, each of which must be prohibited from re-initiating DNA replication within every cell cycle. How cells prevent re-initiation has been a long-standing question in cell biology. In several eukaryotes, cyclin-dependent kinases (CDKs) have been implicated in promoting the block to re-initiation¹, but exactly how they perform this function is unclear. Here we show that B-type CDKs in *Saccharomyces cerevisiae* prevent re-initiation through multiple overlapping mechanisms, including phosphorylation of the origin recognition complex (ORC), downregulation of Cdc6 activity, and nuclear exclusion of the Mcm2-7 complex. Only when all three inhibitory pathways are disrupted do origins re-initiate DNA replication in G2/M cells. These studies show that each of these three independent mechanisms of regulation is functionally important.

The mechanism of eukaryotic replication initiation and the role of CDKs in its regulation have been most extensively characterized in the budding yeast *S. cerevisiae* (reviewed in ref. 1). Initiation events at yeast origins can be divided into two fundamental stages: the assembly of pre-replicative complexes (pre-RCs) and the triggering of new DNA synthesis. The assembly of pre-RCs occurs shortly after mitosis and renders origins competent to initiate DNA synthesis. During this assembly ORC, which binds origins throughout the cell cycle, is joined by additional initiator proteins, including Cdc6 and the Mcm2-7 complex. Passage through the G1 commitment point (Start) then activates the kinases Cdc7–Dbf4 and the B-type CDKs Clb–Cdc28, which together trigger origin unwinding, assembly of the replication fork machinery, initiation of daughter strand synthesis, and pre-RC disassembly.

In addition to triggering initiation, Clb–Cdc28 prevents re-initiation, in part by blocking re-assembly of pre-RCs¹. This block

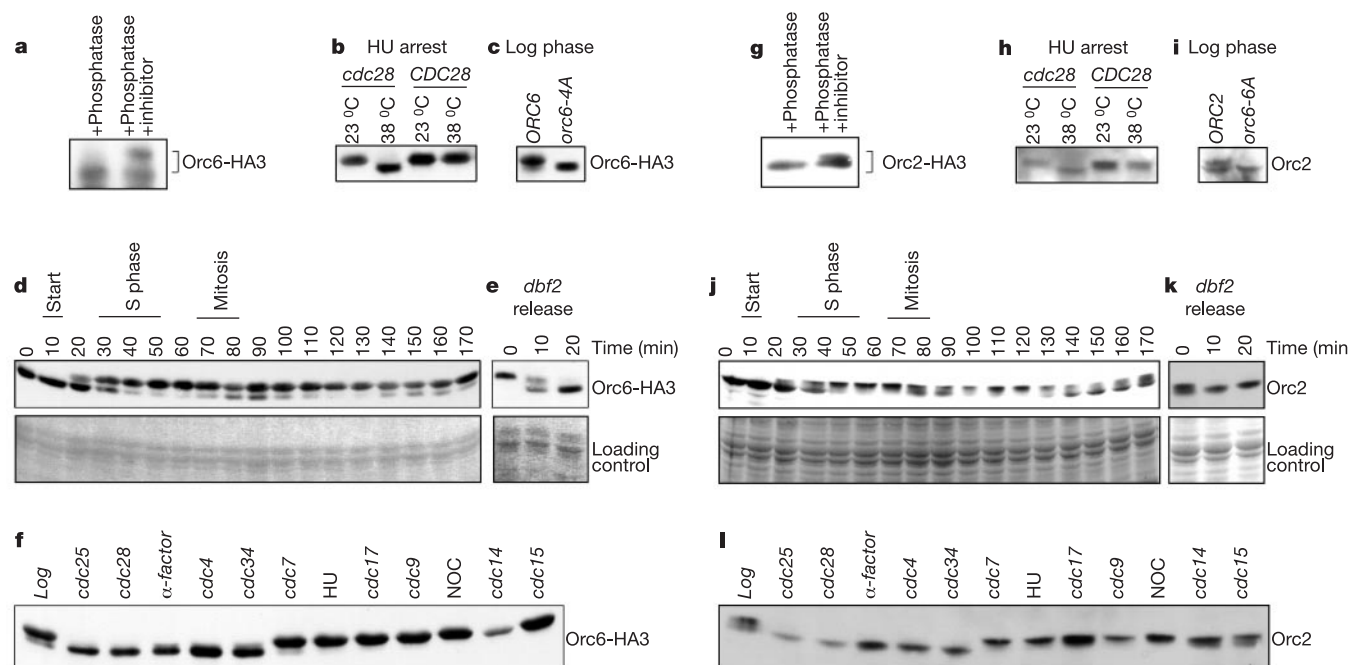


Figure 1 Clb–Cdc28 phosphorylation of Orc6 and Orc2 *in vivo*. **a–f**, Immunoblot of Orc6 using anti-haemagglutinin (HA) detection of Orc6-HA3 (**a–d**, **f**) or anti-Orc6 (**e**). **g–l**, Immunoblot of Orc2 using anti-HA detection of Orc2-HA3 (**g**) or anti-Orc2 (**h–l**). Extracts used in **h**, **j–l** were identical to those used in **b**, **d–f**, respectively. **a**, **g**, Immunoprecipitates from YJL921 (*ORC6-HA3*) (**a**) or YJL963 (*ORC2-HA3*) (**g**) treated with λ -phosphatase with or without phosphatase inhibitors. **b**, **h**, YJL934 (*cdc28-4 ORC6-HA3*) or YJL865 (*CDC28 ORC6-HA3*) grown at 23 °C were arrested in early S phase with hydroxyurea (HU) (after a pre-arrest in G1 with α -factor) then shifted to either 38 °C or kept at 23 °C for a further 3 h. **c**, Log phase YJL865 (*ORC6-HA3*) and YJL1394 (*orc6-4A-HA3*). **i**, Log phase YJL3155 (*ORC2*) and YJL1737 (*orc2-6A*). **d**, **j**, YJL865 (*ORC6-HA3*)

cells were released (time 0) from an α -factor arrest in G1 and samples taken every 10 min for analysis by immunoblot, FACS (to determine time of S phase), and budding index with 4,6-diamidino-2-phenylindole (DAPI) staining (to determine time of Start and mitosis). The 80-min time point in **j** is absent. **e**, **k**, YJL1937 (*dbf2-2 ORC0*) cells were grown at 37 °C for 150 min to arrest them in late mitosis, released from the arrest (time 0) by shifting them to 23 °C, and sampled every 10 min for immunoblot analysis. **f**, **l**, immunoblot of *ORC6-HA3 cdc* strains arrested by growth at 37 °C for 2–3 h (until more than 95% have appropriate bud morphology); drug arrests were performed on YJL864 (*ORC6-HA3*) using α -factor, hydroxyurea or nocodazole (NOC).

is maintained until the kinase is inactivated at the end of mitosis, ensuring that origins initiate only once per cell cycle. CDKs have also been implicated in preventing re-replication in *Schizosaccharomyces pombe*, *Drosophila melanogaster* and *Xenopus laevis*¹; however, the mechanism by which CDKs prevent pre-RC re-assembly and the identity of their relevant inhibitory targets are poorly understood. Clb–Cdc28 reduces Cdc6 levels through phosphorylation of Cdc6, which promotes its ubiquitin-mediated degradation^{2–4}, and phosphorylation of the transcriptional activator Swi5, which prevents it from entering the nucleus and inducing Cdc6 expression⁵. Clb–Cdc28 also promotes the net nuclear export of MCM proteins, leading to their exclusion from the nucleus in G2 and M phases^{6,7}. However, constitutive expression of stabilized or non-phosphorylatable Cdc6 (ref. 2; and data not shown) or constitutive nuclear localization of Mcm2–7 (ref. 6) do not induce re-replication within a cell cycle. Therefore, it has not been possible to establish the functional importance of these mechanisms in the block to re-replication. Moreover, these results leave open the possibility that Clb–Cdc28 targets additional replication proteins to maintain this block.

Here we have examined the possible regulation of ORC by Clb–Cdc28 and its relevance to the control of replication. Three of the six ORC proteins, Orc1, Orc2 and Orc6, have consensus CDK phosphorylation sites ((S/T)-P-X-(K/R)), indicating that they might be phosphorylated by the kinase *in vivo*. Consistent with this

possibility, Orc6 and Orc2 each migrated on SDS–polyacrylamide gel electrophoresis (PAGE) as a doublet, which was converted to the faster-migrating form on phosphatase treatment (Fig. 1a, g). The presence of the slower-migrating hyperphosphorylated form was dependent on both Cdc28 (Fig. 1b, h) and the CDK consensus phosphorylation sites (Fig. 1c, i), and was cell-cycle regulated (Fig. 1d, j). Both Orc6 and Orc2 were hypophosphorylated in G1, became hyperphosphorylated after Start, and remained hyperphosphorylated until the next G1 phase. Incomplete conversion to the hypophosphorylated form in the second and third G1 phases was probably due to loss of cell synchrony, as both proteins showed rapid and complete conversion to the hypophosphorylated form in cells synchronously released into G1 phase from a *dbf2* late mitotic arrest (Fig. 1e, k). These findings indicate that Cdc28 phosphorylates Orc6 and Orc2 on at least some of their CDK consensus sites *in vivo*.

We determined more precisely the timing of this phosphorylation by examining Orc6 and Orc2 at different cell-cycle arrests (Fig. 1f, l). Both proteins were hypophosphorylated after Start on *cdc4* or *cdc34* arrest, when Cln–Cdc28 is active⁸, and only became hyperphosphorylated later in G1 at a *cdc7* arrest, when Clb–Cdc28 is active⁸. Orc2 and Orc6 remained hyperphosphorylated at all arrest points later in the cell cycle, matching the persistence of Clb–Cdc28 kinase activity through late anaphase. These data indicate that Clb–Cdc28 and not Cln–Cdc28 is responsible for the cell-cycle-regulated

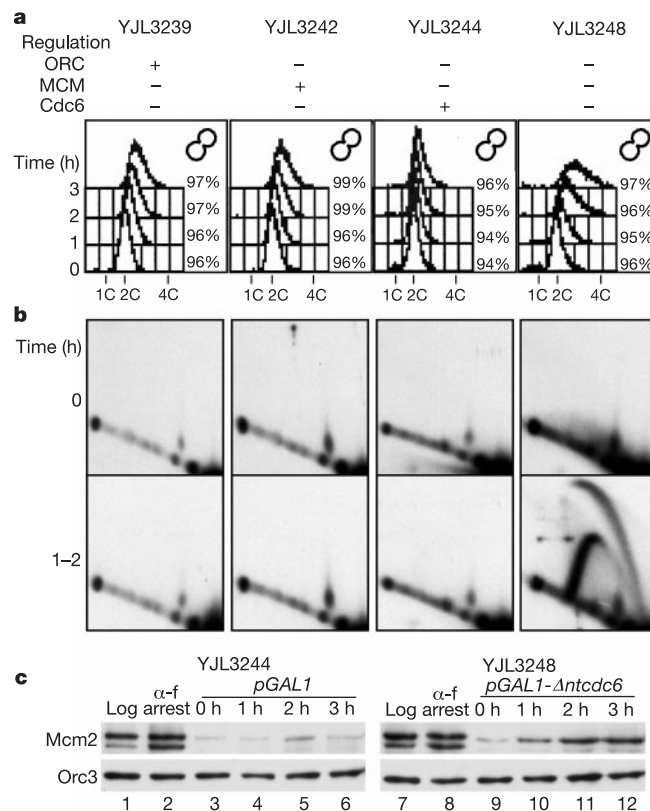


Figure 2 Induction of re-replication and re-initiation in G2/M by deregulation of ORC, Mcm2–7 and Cdc6. ORC phosphorylation, Mcm2–7 localization and Cdc6 expression were deregulated in YJL3239, YJL3242, YJL3244 and YJL3248, as described in the text. Minus, deregulated; plus, regulated. Deregulation of Cdc6 was conditional and dependent on galactose induction of *pGAL1-Δntcdc6*. Cells were initially grown in medium lacking methionine and containing raffinose to prevent expression of *Δntcdc6*. They were arrested in G2/M by addition of 2 mM methionine to induce depletion of Cdc20 followed 2.5 h later by 15 μg ml^{–1} nocodazole. After a further 30 min, galactose was added to induce *Δntcdc6* at 0 h. **a**, Budding index and flow cytometry. **b**, Strains were sampled at 0 h or between 1 and 2 h (sampled every 2 min and pooled) for analysis of DNA replication

intermediates at *ARS305* by neutral–neutral two-dimensional gel electrophoresis. Similar two-dimensional gel results obtained from cells sampled between 0 and 1 h, 1 and 2 h, and 2 and 3 h were seen with congenic strains that were wild type for *CDC20* and arrested in G2/M solely with nocodazole (data not shown). No replication intermediates were induced in any strain if dextrose instead of galactose was added to the medium to repress the *GAL1* promoter. **c**, Mcm2 reassociates with chromatin during re-replication. Re-replication was induced in YJL3244 and YJL3248 as described above, and chromatin-enriched fractions²⁰ were analysed at the indicated time points (lanes 3–6, 9–12) by immunoblotting with anti-Mcm2 and anti-Orc3 antibodies. Log phase (lanes 1, 7) and α-factor (α-f; lanes 2, 8) arrested cells were examined together.

hyperphosphorylation of Orc2 and Orc6. Moreover, the results demonstrate that this phosphorylation is independent of Cdc7–Dbf4.

To determine the function of Orc2 and Orc6 phosphorylation, we constructed a strain in which the phospho-acceptor residues of all CDK consensus sites in these proteins were mutated to alanine. The strain was indistinguishable from its congenic wild-type parent in growth rate, plasmid loss rates (a measure of the efficiency of replication initiation), and flow cytometry profile (data not shown). Similar results were obtained with a strain in which the last remaining CDK consensus site in ORC (on Orc1) was also mutated (data not shown). Thus, phosphorylation of ORC on its CDK consensus sites, like the reduction of Cdc6 levels and the nuclear exclusion of Mcm2–7, is not essential for the block to re-initiation. These experiments also show that phosphorylation of ORC proteins on their CDK consensus sites is not required for the initiation of DNA replication.

To test whether Clb–Cdc28 uses several overlapping mechanisms as a safeguard against re-initiation, we constructed strains that combined various disruptions of ORC, Cdc6 and Mcm2–7 regulation. Phosphorylation of Orc2 and Orc6 on their consensus CDK sites was eliminated by mutating these sites as described above. Nuclear exclusion of Mcm2–7 by Clb–Cdc28 was disrupted by fusing two tandem copies of the SV40 nuclear localization signal (NLS) onto Mcm7 (ref. 6). In both cases, the wild-type genes were precisely replaced by their mutant counterpart. The restriction of Cdc6 expression to G1 phase was overridden by expressing a partially stabilized form of Cdc6, Δ ntCdc6, under the control of the galactose-inducible *GAL1* promoter. Δ ntCdc6 contains an amino-terminal truncation of amino acids 2–46, which removes

sequences that facilitate Cdc6 degradation² and are necessary for Clb–Cdc28 association⁹. Despite these mutations, Δ ntCdc6 can fully substitute for wild-type Cdc6 (as measured by plasmid loss rates) when expressed from the *CDC6* promoter (data not shown). *pGAL1- Δ ntCdc6* was introduced in addition to the endogenous *CDC6* gene.

We constructed three congenic strains (YJL3239, YJL3242 and YJL3244) containing all three possible pairwise combinations of regulatory perturbations described above and examined them for the ability to re-replicate their DNA at a G2/M phase arrest (when replication is complete and Clb–Cdc28 kinase activity is high). To achieve a tight arrest, cells were both depleted of Cdc20, which is required for the metaphase–anaphase transition¹⁰, and exposed to nocodazole, a microtubule-destabilizing agent that disrupts mitotic spindles. Only after cells were arrested was Δ ntCdc6 induced by galactose. None of these strains increased their DNA content significantly beyond 2C (as measured by flow cytometry, Fig. 2a) or displayed any actively replicating chromosomes (Supplementary Information Fig. 1), which migrate with retarded mobility during pulsed-field gel electrophoresis (PFGE)¹¹. Moreover, *ARS305* (Fig. 2b) and *ARS1* (data not shown), which normally fire in early and early to mid S phase, respectively^{12,13}, showed no signs of re-initiation or passive re-replication by neutral–neutral two-dimensional gel electrophoresis. We conclude that the block to re-initiation in G2/M phase remains largely intact despite simultaneous disruption of any two of the three regulatory mechanisms described above.

We next tested whether a strain containing disruptions in all three regulatory mechanisms (YJL3248) would undergo re-replication at a G2/M phase arrest. In contrast to the results above, galactose induction of Δ ntCdc6 resulted in an increase in DNA content from 2C to ~3C (Fig. 2a) and induction of initiation bubbles at *ARS305*, *ARS121*, *ARS607* (Fig. 2b; see also Supplementary Information Fig. 2b, e, f) and *ARS1* (data not shown). The re-initiation at *ARS305* was dependent on its *ARS* consensus sequence and ORC

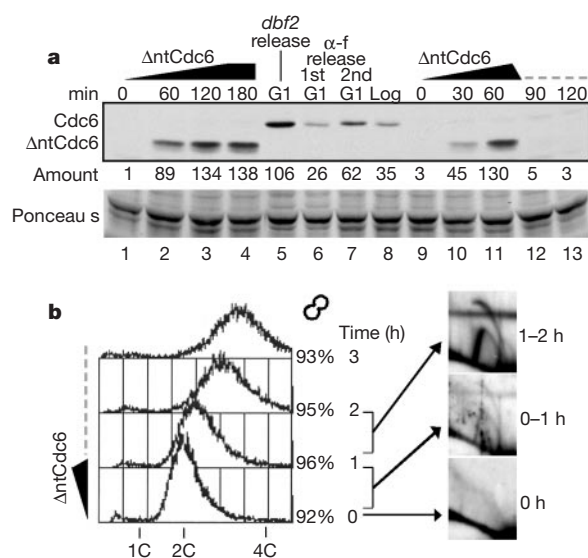


Figure 3 Re-replication and re-initiation are induced with physiological levels of Δ ntCdc6. **a**, Immunoblot with anti-Cdc6 antibodies of re-replicating strain YJL3248 (lanes 1–4, 9–13) and *dbf2-2* strain YJL1937 (lanes 5–8). Δ ntCdc6 was induced in YJL3248 as described in Fig. 2 (lanes 1–4) or as in Fig. 2 with the addition of dextrose after 60 min to repress further induction (lanes 9–13). Endogenous levels of Cdc6 in YJL1937 were monitored every 10 min after release from a *dbf2-2* late mitotic arrest or every 15 min after release from an α -factor (α -f) arrest. Time points containing the peak levels immediately after *dbf2* release (lane 5, G1) or in the first (lane 6, 1st G1) or second (lane 7, 2nd G1) G1 phases after α -factor release are shown and compared to a log phase population (lane 8). Peak G1 levels in cycling cells seem to be best represented by peak levels after *dbf2* release. Band intensities quantified by densitometry are expressed in arbitrary units (amount). **b**, FACS analysis and budding indices of YJL3248 at indicated times during a transient 1 h induction of Δ ntCdc6 as described in **a** (lanes 9–13); two-dimensional gel analysis of *ARS607* taken at 0, 0–1 and 1–2 h (the latter two sampled every 2 min and pooled).

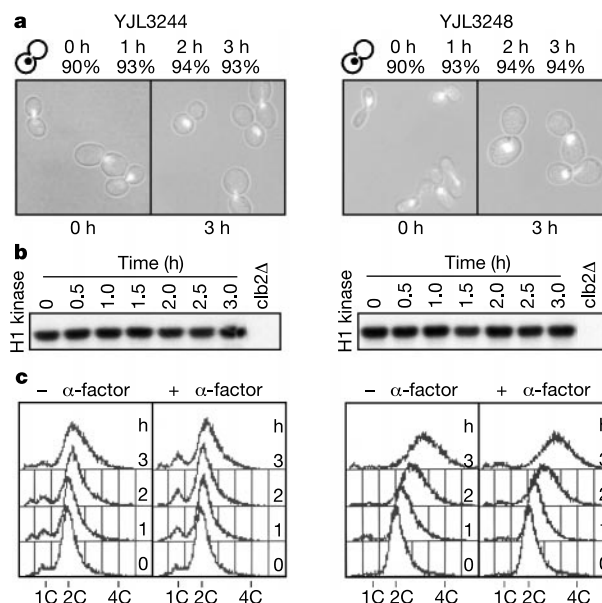


Figure 4 Re-replicating cells remain in G2/M phase. YJL3244 and YJL3248 are a control and re-replicating strain, respectively, induced to re-replicate as described in Fig. 2. **a**, Percentage of cells that are large-budded with a single nucleus. Pictures show DAPI fluorescence overlaying bright field microscopy of cells at 0 and 3-h time points. **b**, H1 kinase assays performed on anti-Clb2 immunoprecipitates taken every 0.5 h after galactose induction of re-replication. Measurements were performed in the linear range of the assay. **c**, Galactose induction of re-replication was performed in the presence or absence of α -factor (to arrest any cells entering G1 phase) and samples were taken every hour for flow cytometry.

binding site¹⁴ (Supplementary Information Fig. 2a), suggesting that re-initiation occurred through the same ORC-dependent mechanism as S-phase initiation. Re-replication was also accompanied by re-association of Mcm2 with chromatin (Fig. 2c), suggesting that MCM complexes reloaded onto origins to re-initiate replication. Furthermore, the mobility of all chromosomes was retarded during PFGE (Supplementary Information Fig. 1a), with Southern analysis confirming that both large (chromosomes 4 and 7, Supplementary Information Fig. 1b) and small (chromosome 3; data not shown) chromosomes experienced difficulty entering the gel. Hence, all chromosomes seemed to participate in the re-replication. Together our results indicate that ORC, Cdc6 and Mcm2-7 must be simultaneously deregulated to re-initiate DNA replication.

Given that re-initiation arose after induction of a partially stabilized form of Cdc6 from the strong *GAL1* promoter, we wished to confirm that the re-initiation was due to ectopic expression and not massive overexpression of Δ ntCdc6 (which might cause additional, unknown perturbations). Western analysis indicated that the level of Δ ntCdc6 induced in our re-replicating strain after 2 h in galactose was less than twofold higher than peak levels of endogenous Cdc6 expressed in early G1 phase (Fig. 3a). We also induced Δ ntCdc6 synthesis for only 1 h to mimic the normally transient G1 expression of Cdc6. Considerable re-replication and re-initiation was still observed after this transient induction, and some re-initiation was detected during the induction before Δ ntCdc6 had fully accumulated (Fig. 3b). Induction of the fully unstable wild-type Cdc6 in G2/M also triggered re-replication (Supplementary Information Fig. 3). Thus, ectopic expression of Δ ntCdc6 without any significant overexpression was sufficient to induce re-replication in the triply deregulated strain.

Despite re-initiating DNA replication, the triply deregulated strain did not completely duplicate its DNA, suggesting that we had not removed all restraints on re-replication. Although several origins re-initiated efficiently, some origins did not, including two early (*ARS306* and *ARS307*; Supplementary Information Fig. 2c, d) and two late (*ARS501* and *ARS1413*; Supplementary Information

Fig. 2h, i) origins. Two-dimensional gel analysis of *ARS305*, *ARS306* and *ARS307* demonstrated that they all initiated efficiently during the S phase preceding the induced re-replication (data not shown). These observations suggest that additional mechanisms prevent re-initiation of some origins and raise the question of what distinguishes these origins from those that do re-initiate. Y arcs were clearly induced at origins that failed to re-initiate, indicating that these origins were still passively re-replicated, presumably by replication forks that had re-initiated from neighbouring origins. The weaker intensity of these arcs, however, suggests that re-elongation may also have been partially inhibited. Forks originating from *ARS305* and *ARS607* seemed to have some difficulty re-replicating fragments only 30–35 kilobases (kb) away (see Supplementary Information Fig. 2c, g), even though replication forks can travel at least 100–200 kb in S phase¹². A more quantitative genome-wide analysis of re-replication will be needed to confirm and fully characterize any remaining inhibition of re-initiation and/or re-elongation in our triply deregulated strain. Nonetheless, additional mechanisms besides those we specifically deregulated are likely to restrict re-replication within a single cell cycle. Some of these mechanisms may provide further means of inhibiting ORC, Cdc6 or Mcm2-7 (just as Cdc18 in *S. pombe* is independently restrained by both decreased expression and phosphorylation¹⁵), whereas others may target additional replication proteins.

Unscheduled DNA replication has also been reported to arise from transient inactivation of CDK activity in G2 or G2/M-arrested cells^{16–19}. This CDK inactivation, however, resets the cell cycle to G1 phase in the absence of mitosis, effectively inducing S phase in a new cell cycle and not re-initiation within G2 or G2/M phase of the original cell cycle. In *S. cerevisiae*, for example, transient inactivation of Clb–Cdc28 in G2/M-arrested cells, by overexpression of the CDK inhibitor Sic1, induces transcription of G1-specific genes and triggers the G1-specific event of budding¹⁷. Consistent with the cell cycle being reset to G1 phase, the new round of replication that ensues on release of Clb–Cdc28 inactivation can be blocked by the mating pheromone α -factor which arrests cells in G1 (J. Diffley,

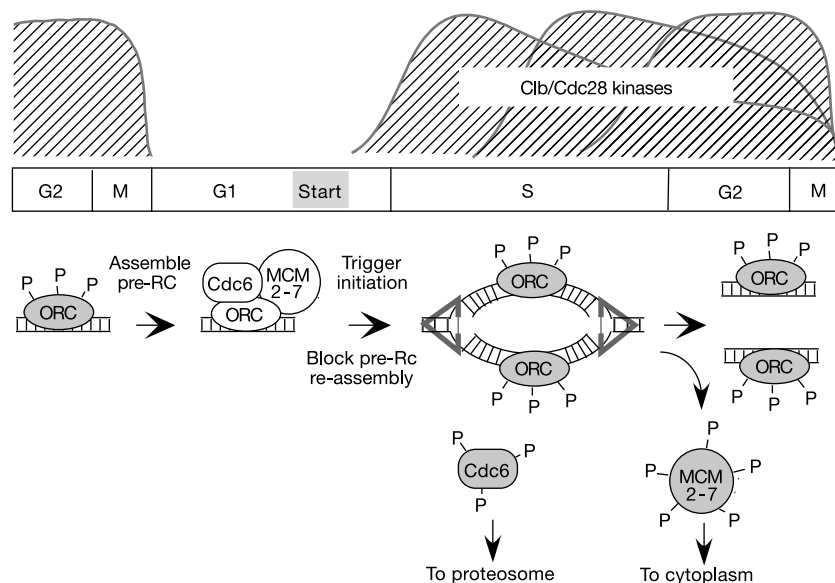


Figure 5 Model for Clb–Cdc28 inhibition of re-replication through several overlapping mechanisms. Mcm2-7 and Cdc6 join ORC at the origin to form the pre-replicative complex (pre-RC) in G1 phase when Cdc28 kinase activity is low. Induction of Clb–Cdc28 after Start helps to trigger initiation, resulting in assembly of the replication fork machinery (triangles) and disassembly of the pre-RC. The kinase simultaneously prevents re-initiation by at least three overlapping mechanisms: (1) phosphorylating Cdc6 and facilitating its polyubiquitination and degradation^{3,9}; (2) promoting nuclear exclusion of

MCM proteins^{6,7}, most likely by phosphorylating the MCM proteins (A. Rosales and J.J.L., unpublished data); and (3) inhibiting ORC function through phosphorylation. No single mechanism is individually essential to prevent re-replication in G2/M cells, as each is sufficient to maintain this block. Additional replication proteins involved in pre-RC assembly or triggering (not shown) may also be inhibited by Clb–Cdc28.

personal communication). Thus, although this unscheduled DNA replication confirms that cell-cycle position is determined by CDK activity¹⁸, it does not address how this CDK activity prevents re-replication within a cell cycle.

Our triply deregulated strain did not bypass mitosis and enter a G1-like state before they re-replicated. The cells did not rebud (Figs 2a and 4a), maintained high mitotic levels of Clb2–Cdc28 kinase activity (Fig. 4b), and were able to re-replicate in the presence of α -factor (Fig. 4c). They also did not break through the mitotic arrest and enter S phase of the next cell cycle, as cells maintained a single, undivided nucleus while they re-replicated (Fig. 4a). Furthermore, by inducing Δ ntCdc6 rather than the full-length protein, we avoided Cdc6 association with, and possible inhibition of, Clb–Cdc28 (ref. 9). We conclude that the block to re-initiation was removed in the triply deregulated strain because we had simultaneously rendered ORC, Cdc6 and Mcm2-7 refractory to the inhibitory action of Clb–Cdc28, and not because we had inadvertently inactivated Clb–Cdc28.

These findings indicate that Clb–Cdc28 uses at least three overlapping inhibitory pathways involving phosphorylation of ORC, decreased expression of Cdc6, and nuclear exclusion of Mcm2-7 to prohibit re-initiation of DNA replication in budding yeast (Fig. 5). Because any one of these mechanisms is sufficient to block re-initiation in the absence of the others (Figs 2 and 3, strains YJL3239, YJL3242 and YJL3244) no single mechanism is individually essential for this block. We propose that budding yeast uses a combination of overlapping mechanisms targeting distinct initiation proteins to ensure that none of its hundreds of replication origins re-initiate within a cell cycle. Although no overt re-initiation was seen when any single mechanism was disrupted, each mechanism may be important for preserving long-term genome stability by keeping the frequency of re-initiation events extremely low over the course of many cell divisions. Hence these overlapping mechanisms should be considered mutually reinforcing and not necessarily redundant.

Although multiple mechanisms targeting distinct proteins must be disrupted to induce re-initiation, given that these proteins work together in a complex, it is conceivable that mutation or perturbation of one of these proteins could override enough mechanisms to trigger re-initiation. This may account for reports of re-replication arising from perturbation of just *CDC6* or its *S. pombe* orthologue *CDC18*. In one report, *cdc6-3* (a mutant severely defective for initiation) accumulated a 2.5–3C DNA content in G2/M phase and exhibited both persistent initiation intermediates and persistent association of MCM proteins with chromatin²⁰. Although there was no direct demonstration that these persistent signs of replication arose from re-initiation, it is possible that the *cdc6-3* mutation induced re-initiation by counteracting mechanisms inhibiting ORC and MCM proteins as well as those inhibiting Cdc6. Similarly, whereas massive overexpression of Cdc18 is sufficient to induce repeated rounds of re-replication in *S. pombe*^{21,22}, more modest overexpression does not do so unless a second initiation protein, Cdt1, is simultaneously overexpressed²³ or CDK phosphorylation of Orc2 is prevented²⁴. These results suggest that, in addition to overriding the G1-specific expression of Cdc18, massive overexpression of Cdc18 may overwhelm other mechanisms that prevent re-replication in *S. pombe*.

CDKs in other eukaryotes may also depend on several downstream inhibitory targets to prevent re-replication within a single cell cycle. Such a model could help explain why disruption of Cdc6 regulation in humans²⁵ and *Xenopus*²⁶ is not sufficient to induce re-replication. Basic mechanistic strategies used by *S. cerevisiae* to prevent re-replication, such as destruction, re-localization, or modification of replication proteins, are likely to be conserved. The precise implementation of those strategies, however, may vary. For example, although in most eukaryotes MCM proteins are not excluded from the nucleus in S and G2 phase¹, in mammalian cells Cdc6 is excluded after G1 phase^{27–30}. The relative importance of

various strategies may also differ in different organisms and at different times in the cell cycle. Nonetheless, by using multiple inhibitory mechanisms to target more than one replication protein, eukaryotic cells can maintain a tight block to re-replication at the hundreds to thousands of replication origins in their genomes. □

Methods

Details of plasmid/strain construction and experimental assays can be found in Supplementary Information. Yeast growth, galactose induction, methionine repression and cell cycle arrest/release were performed as described⁶.

The strains with deregulated ORC phosphorylation, Mcm2-7 localization and Cdc6 expression (Fig. 2) were: YJL3239 (*ORC2 ORC6 MCM7-2NLS CDC6 pGAL1- Δ ntcdc6 pMET3-CDC20*); YJL3242 (*orc2-6A orc6-4A MCM7-2nls3A CDC6 pGAL1- Δ ntcdc6 pMET3-CDC20*); YJL3244 (*orc2-6A orc6-4A MCM7-2NLS CDC6 pGAL1 pMET3-CDC20*); and YJL3248 (*orc2-6A orc6-4A MCM7-2NLS CDC6 pGAL1- Δ ntcdc6 pMET3-CDC20*).

Received 9 April; accepted 31 May 2001.

- Kelly, T. J. & Brown, G. W. Regulation of chromosome replication. *Annu. Rev. Biochem.* **69**, 829–880 (2000).
- Drury, L. S., Perkins, G. & Diffley, J. F. The Cdc4/34/53 pathway targets Cdc6p for proteolysis in budding yeast. *EMBO J.* **16**, 5966–5976 (1997).
- Drury, L. S., Perkins, G. & Diffley, J. F. The cyclin-dependent kinase Cdc28p regulates distinct modes of Cdc6p proteolysis during the budding yeast cell cycle. *Curr. Biol.* **10**, 231–240 (2000).
- Ellasser, S., Chi, Y., Yang, P. & Campbell, J. L. Phosphorylation controls timing of Cdc6p destruction: a biochemical analysis. *Mol. Biol. Cell* **10**, 3263–3277 (1999).
- Moll, T., Tebb, G., Surana, U., Robitsch, H. & Nasmyth, K. The role of phosphorylation and the CDC28 protein kinase in cell cycle-regulated nuclear import of the *S. cerevisiae* transcription factor SWI5. *Cell* **66**, 743–758 (1991).
- Nguyen, V. Q., Co, C., Irie, K. & Li, J. J. Clb/Cdc28 kinases promote nuclear export of the replication initiator proteins Mcm2-7. *Curr. Biol.* **10**, 195–205 (2000).
- Labib, K., Diffley, J. F. & Kearsey, S. E. G1-phase and B-type cyclins exclude the DNA-replication factor Mcm4 from the nucleus. *Nature Cell Biol.* **1**, 415–422 (1999).
- Schwob, E., Bohm, T., Mendenhall, M. D. & Nasmyth, K. The B-type cyclin kinase inhibitor p40SIC1 controls the G1 to S transition in *S. cerevisiae*. *Cell* **79**, 233–244 (1994).
- Ellasser, S., Lou, F., Wang, B., Campbell, J. L. & Jong, A. Interaction between yeast Cdc6 protein and B-type cyclin/Cdc28 kinases. *Mol. Biol. Cell* **7**, 1723–1735 (1996).
- Visintin, R., Prinz, S. & Amon, A. CDC20 and CDH1: a family of substrate-specific activators of APC-dependent proteolysis. *Science* **278**, 460–463 (1997).
- Hennessy, K. M., Clark, C. D. & Botstein, D. Subcellular localization of yeast *CDC46* varies with the cell cycle. *Genes Dev.* **4**, 2252–2263 (1990).
- Newlon, C. S. *et al.* Analysis of replication origin function on chromosome III of *Saccharomyces cerevisiae*. *Cold Spring Harb. Symp. Quant. Biol.* **58**, 415–423 (1993).
- Brewer, B. J. *et al.* The topography of chromosome replication in yeast. *Cold Spring Harb. Symp. Quant. Biol.* **58**, 425–434 (1993).
- Bell, S. P. & Stillman, R. ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. *Nature* **357**, 128–134 (1992).
- Jallepalli, P. V., Brown, G. W., Muzi-Falconi, M., Tien, D. & Kelly, T. J. Regulation of the replication initiator protein p55cdc18 by CDK phosphorylation. *Genes Dev.* **11**, 2767–2779 (1997).
- Sauer, K., Knoblich, J. A., Richardson, H. & Lehner, C. F. Distinct modes of cyclin E/cdc2c kinase regulation and S-phase control in mitotic and endoreduplication cycles of *Drosophila* embryogenesis. *Genes Dev.* **9**, 1327–1339 (1995).
- Dahmann, C., Diffley, J. F. & Nasmyth, K. A. S-phase-promoting cyclin-dependent kinases prevent re-replication by inhibiting the transition of replication origins to a pre-replicative state. *Curr. Biol.* **5**, 1257–1269 (1995).
- Broek, D., Bartlett, R., Crawford, K. & Nurse, P. Involvement of p34cdc2 in establishing the dependency of S phase on mitosis. *Nature* **349**, 388–393 (1991).
- Noton, E. & Diffley, J. F. CDK inactivation is the only essential function of the APC/C and the mitotic exit network proteins for origin resetting during mitosis. *Mol. Cell* **5**, 85–95 (2000).
- Liang, C. & Stillman, R. Persistent initiation of DNA replication and chromatin-bound MCM proteins during the cell cycle in cdc6 mutants. *Genes Dev.* **11**, 3375–3386 (1997).
- Muzi-Falconi, M., Brown, G. & Kelly, T. J. cdc18+ regulates initiation of DNA replication in *Schizosaccharomyces pombe*. *Proc. Natl Acad. Sci. USA* **93**, 1666–1670 (1996).
- Nishitani, H. & Nurse, P. p55cdc18 plays a major role controlling the initiation of DNA replication in fission yeast. *Cell* **83**, 397–405 (1995).
- Nishitani, H., Lygerou, Z., Nishimoto, T. & Nurse, P. The Cdt1 protein is required to license DNA for replication in fission yeast. *Nature* **404**, 2758–2767 (2000).
- Vas, A., Mok, W. & Leatherwood, J. Control of DNA re-replication via Cdc2 phosphorylation sites in ORC. *Mol. Cell. Biol.* (2001) (in the press).
- Petersen, B. O. *et al.* Cell cycle- and cell growth-regulated proteolysis of mammalian CDC6 is dependent on APC-CDH1. *Genes Dev.* **14**, 2330–2343 (2000).
- Pelizon, C., Madine, M. A., Romanowski, P. & Laskey, R. A. Unphosphorylatable mutants of cdc6 disrupt its nuclear export but still support DNA replication once per cell cycle. *Genes Dev.* **14**, 2526–2533 (2000).
- Jiang, W., Wells, N. J. & Hunter, T. Multistep regulation of DNA replication by Cdk phosphorylation of HsCdc6. *Proc. Natl Acad. Sci. USA* **96**, 6193–6198 (1999).
- Petersen, B. O., Lukas, J., Sorensen, C. S., Bartek, J. & Helin, K. Phosphorylation of mammalian CDC6 by cyclin A/CDK2 regulates its subcellular localization. *EMBO J.* **18**, 396–410 (1999).
- Saha, P. *et al.* Human CDC6/Cdc18 associates with Orc1 and cyclin-cdk and is selectively eliminated from the nucleus at the onset of S phase. *Mol. Cell. Biol.* **18**, 2758–2767 (1998).
- Fujita, M. *et al.* Cell cycle regulation of human CDC6 protein. Intracellular localization, interaction with the human mcm complex, and CDC2 kinase-mediated hyperphosphorylation. *J. Biol. Chem.* **274**, 25927–25932 (1999).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We thank T. Wang, S. Chu and J. Whangbo for early observations about Orc2 and Orc6 phosphorylation; U. Tran for analysis of Δ ntCdc6 plasmid loss rates; L. Huang and J. Gitschier for aid with the PFGE analysis; B. Stillman, C. Liang, S. Bell, J. Ubersax, A. Rudner, M. K. Raghuraman, F. Uhlman, D. Morgan and D. Toczyski for advice and reagents; and A. Sil, I. Herskowitz, A. Johnson, D. Morgan, P. O'Farrell, E. Blackburn, D. Toczyski, B. Thorner, J. Ubersax and C. Takizawa for critical reading of the manuscript. This work was supported by ACS and NIH (J.J.L.) and an NIH training grant (V.Q.N.). Part of this work was performed when J.J.L. was a Markey, Searle and Rita Allen Foundation Scholar.

Correspondence and requests for materials should be addressed to J.J.L. (e-mail: jli@itsa.ucsf.edu).

Multiple pathways cooperate in the suppression of genome instability in *Saccharomyces cerevisiae*

Kyungjae Myung, Clark Chen & Richard D. Kolodner

Ludwig Institute for Cancer Research, Cancer Center and Department of Medicine, University of California San Diego School of Medicine, La Jolla, California 92093, USA

Gross chromosome rearrangements (GCRs), such as translocations, deletion of a chromosome arm, interstitial deletions and inversions, are often observed in cancer cells^{1–3}. Spontaneous GCRs are rare in *Saccharomyces cerevisiae*; however, the existence of mutator mutants with increased genome instability suggests that GCRs are actively suppressed^{4,5}. Here we show by genetic analysis that these genome rearrangements probably result from DNA replication errors and are suppressed by at least three interacting pathways or groups of proteins: S-phase checkpoint functions⁶, recombination proteins⁴ and proteins that prevent *de novo* addition of telomeres at double-strand breaks (DSBs). Mutations that inactivate these pathways cause high rates of GCRs and show synergistic interactions, indicating that the pathways that suppress GCRs all compete for the same DNA substrates.

The pathways that suppress GCRs have not been well characterized. Using a mutator assay that detects GCRs in *S. cerevisiae*, we demonstrated that mutations in some DNA repair genes and genes encoding S-phase checkpoint functions cause increased rates of accumulating GCRs^{4,5}. Three types of GCRs were seen including translocations, deletions and terminal deletion of the ends of chromosomes with addition of a new telomere (telomere additions). All of the mutator mutants had increased rates of telomere additions, whereas only a subset had increased rates of translocations, suggesting that telomerase activity has a role in the formation of GCRs. Thus, mutations in genes that affect telomerase activity were tested for their effect on GCR rates (Table 1). Mutations that inactivated the catalytic activity of telomerase (*est1*, *est2*, *est3*, *cdc13-2*, *tlc1*) or inactivated proteins that affect telomeres (*tel1*, *sir1-4*, *yku70*, *yku80*, *rif1*, *rif2*)^{6,7} had no significant effect on the GCR rate. The *stn1-13* mutation, which causes longer telomeres⁸, resulted in a small increase in the GCR rate. A *pif1*Δ deletion mutation caused an approximately 1,000-fold increase in the GCR rate. PIF1 is a DNA helicase that functions in both mitochondria and the nucleus, and *pif1* mutations result in both longer telomeres and loss of mitochondrial function^{9–11}. The *pif1-m2* allele, which only inactivates the nuclear function of PIF1, increased the GCR rate, whereas the *pif1-*

m1 allele, which only inactivates the PIF1 mitochondrial function, had no effect. This indicates that the nuclear PIF1 has a principal function in suppression of GCRs.

PIF1 is an inhibitor of telomerase that suppresses telomere additions at HO-induced DSBs by about 14-fold^{10,11}. The GCRs observed in the *pif1-m2* mutant were telomere additions (Table 2b). Mutations that inactivated the catalytic activity of telomerase (*est1*, *est2*, *est3*, *cdc13-2*, *tlc1*) reduced the GCR rate caused by the *pif1-m2* mutation to wild-type levels (Table 1), indicating that telomere additions in *pif1* mutants require telomerase. A *tel1* mutation had no effect on the GCR rate caused by the *pif1-m2* mutation, and 80% of the GCRs seen in the double mutant were telomere additions (Table 2b). This indicates that there is sufficient telomerase activity in a *tel1 pif1* double mutant for *de novo* telomere addition to occur. Mutations in *RIF1* or *RIF2*, which encode proteins that interact with RAP1 (refs 6 and 7), had no effect on the GCR rate caused by the *pif1-m2* mutation. In contrast, a *rif2* mutation increased the length of telomeres added at an HO break-site adjacent to a 81-base-pair (bp) TG repeat¹², a difference that may reflect the fact that the telomere additions seen here do not occur at long TG repeats^{4,5,13}. Mutations in *yKU70* or *yKU80*, which encode the Ku70/80 heterodimer that binds to DSBs⁶, reduced the GCR rate caused by the *pif1-m2* mutation to almost wild-type levels. Mutations in *SIR2*, *SIR3* or *SIR4*, but not *SIR1*, reduced the GCR rate caused by the *pif1-m2* mutation. The effect of *sir* is probably indirect, as a mutation in *HML*α significantly reduced the defect caused by *sir2* (ref. 14). The *stn1-13* mutation reduced the GCR rate caused by the *pif1-m2* mutation, a result that may reflect the reduced end protection that occurs in both *stn1* and *cdc13* mutants¹⁵. These results indicate that the GCRs that occur in the *pif1-m2* mutant require telomerase and other proteins (Ku70/80, CDC13) that interact with DSBs and effect telomeres.

Mutations that inactivate S-phase checkpoint functions cause an increase in the rate of telomere-addition GCRs⁵. This has been suggested to occur because of errors that normally take place during DNA replication, and S-phase checkpoint defects either cause reduced repair or allow repair to occur in a phase of the cell cycle where telomere addition is the principal outcome^{5,12}. To investigate the role of telomerase in the GCRs resulting from S-phase checkpoint defects, interactions between selected mutations that affect

Table 1 Mutations that alter telomerase activity effect the GCR rate

Relevant genotypes	PIF1		<i>pif1-m2</i>	
	Strain number	Mutation rate (Can ⁺ 5-FOA)	Strain number	Mutation rate (Can ⁺ 5-FOA)
Wild type	3615	3.5×10^{-10} (1)	4343	8.3×10^{-6} (237)
<i>est1</i> Δ	4345	1.5×10^{-10} (0.4)	4365	2.3×10^{-10} (0.7)
<i>est2</i> Δ	4347	1.2×10^{-10} (0.3)	4367	5.8×10^{-10} (1.7)
<i>est3</i> Δ	4349	1.5×10^{-10} (0.4)	4369	2.3×10^{-10} (0.7)
<i>cdc13-2</i>	4351	4.6×10^{-10} (1.3)	4371	4.4×10^{-10} (1.3)
<i>tlc1</i> Δ	4224	3.1×10^{-10} (0.9)	4373	3.3×10^{-10} (0.9)
<i>tel1</i> Δ	3731	2.0×10^{-10} (0.6)	4375	8.2×10^{-8} (234)
<i>sir1</i> Δ	4353	8.9×10^{-10} (2.5)	4377	8.9×10^{-8} (254)
<i>sir2</i> Δ	4355	2.5×10^{-10} (0.7)	4379	6.0×10^{-9} (17)*
<i>sir3</i> Δ	4357	5.0×10^{-10} (1.4)	4381	9.0×10^{-9} (26)
<i>sir4</i> Δ	4359	8.4×10^{-10} (2.4)	4383	3.8×10^{-8} (109)
<i>yku70</i> Δ	3639	4.1×10^{-10} (1)	4385	9.04×10^{-10} (3)
<i>yku80</i> Δ	3640	7.8×10^{-10} (2)	4387	4.0×10^{-9} (11)
<i>rif1</i> Δ	4361	9.9×10^{-10} (3)	4389	7.05×10^{-8} (201)
<i>rif2</i> Δ	4363	5.0×10^{-9} (14)	4391	8.3×10^{-8} (237)
<i>stn1-13</i>	4554	5.3×10^{-9} (15)	4552	2.3×10^{-8} (66)

The GCR rate of a *pif1* deletion strain (RDKY4399) is a 3.53×10^{-7} , which is about four times higher than the GCR rate of the *pif1-m2* strain. Deletion of PIF1 resulted in a petite phenotype owing to the loss of both the nuclear and the mitochondrial function of PIF1. Loss of mitochondrial function of PIF1 did not cause genome instability; a spontaneously arising petite but otherwise wild-type strain (RDKY4397) did not have an increased GCR rate, and a *pif1-m1* strain (RDKY4393) that only lost the mitochondrial function of PIF1 did not have an increased GCR rate. Mutation of *RRM3* (RDKY4395), which encodes another PIF1-like helicase in *S. cerevisiae*, did not affect the GCR rate. Numbers in parentheses indicate the GCR rate relative to the wild-type GCR rate.

* The GCR rate of the *hmlα::Leu2 sir2Δpif1-m2* strain (RDKY4547) was 4.3×10^{-8} (121). Can⁺ 5-FOA is resistant to both canavanine and 5-fluoroorotic acid.

the S-phase checkpoint (*rfc5-1*, *mec1*, *dun1*) and telomerase function (*pif1-m2*, *tlc1*, *est2*, *yku70*) were investigated. Combining the *pif1-m2* mutation with these checkpoint mutations caused a synergistic increase in the GCR rate, with all of the observed GCRs being telomere additions (Table 2). Combining mutations that decrease telomerase activity (*tlc1*, *est2*, *yku70*) with representative checkpoint mutations that cause high GCR rates (*rfc5-1*, *mec1*) caused a decrease of up to 60% in the GCR rate. In the absence of telomerase function, the GCRs resulting from checkpoint defects changed from telomere additions to predominantly translocations and events that fused a large fragment of chromosome V to the end of another chromosome (chromosome fusions). This was particularly marked in the *tlc1 tel1* double mutant. These results are consistent with several conclusions. First, the telomere additions that occur in checkpoint mutants require telomerase activity and the *pif1* mutation makes telomerase activity more available for *de novo* telomere addition. Second, in the absence of telomerase activity, the errors that occur in checkpoint mutants can be repaired by other mutagenic pathways. Finally, telomerase activity is normally limiting so that in checkpoint mutants most errors that occur are not subject to telomere additions but are repaired by other pathways.

Mutations in genes implicated in DSB repair were tested for their effect on the accumulation of GCRs and their interaction with the *pif1-m2* mutation (Table 3). Mutations that inactivated the MRE11–XRS2–RAD50 complex increased the GCR rate, with translocations being the principal class of GCRs observed⁴. Combining the *pif1-m2* mutation with mutations in *MRE11* or *XRS2* had little effect on the GCR rate but shifted the GCR spectrum to telomere additions. The lack of a synergistic interaction with the *pif1-m2* mutation could reflect the fact that *mre11*, *rad50* and *xrs2* mutations cause unusually high GCR rates and affect multiple aspects of DNA metabolism including DNA damage responses and telomerase activity^{5,6,16–18}. A *rad52* mutation caused about a 125-fold increase in the GCR rate, with translocations being the principal class of GCRs observed. Combining the *pif1-m2* mutation with the *rad52* mutation resulted in a synergistic increase in GCR rate, with all of the GCRs observed being telomere additions. Mutations in *RAD51*, *RAD54*, *RAD57*, *RAD59*, *RDH54* and *LIG4* caused much smaller increases in the GCR rate, suggesting that these genes have a small involvement in suppressing GCRs—the

rad57 mutation caused a larger effect at 23 °C, consistent with other studies^{16,19}. Combining mutations in *RAD51* and *RAD59* resulted in a synergistic increase in the GCR rate, suggesting that these genes are redundant^{19–21}. In contrast, the GCR rate of a *rad51 mre11* double mutant was the same as that of a *mre11* mutant (we note that the *mre11* GCR rate is higher than the *rad51 rad59* rate). Combining the *pif1-m2* mutation with these mutations resulted in a synergistic increase in GCR rate (except in the case of *pif1-m2 rdh54*) and in the two cases tested (*pif1-m2 lig4* and *pif1-m2 rdh54*) all of the GCRs were telomere additions. This suggests that although *RAD51*, *RAD54*, *RAD57*, *RAD59*, *RDH54* and *LIG4* individually have only a small involvement in suppressing GCRs, mutations in these genes cause enough of a defect to increase the level of substrates available for telomere additions in the presence of the *pif1-m2* mutation.

The observation that *RAD52* and the *MRE11*–*XRS2*–*RAD50* complex have more important functions in suppressing GCRs than do *RAD51*, *RAD54*, *RAD57*, *RAD59*, *RDH54* and *LIG4* and also the synergistic interaction between *rad51* and *rad59* mutations, is most consistent with the effects of mutations in these genes on break-induced replication^{19,21} and is inconsistent with their effect on classical DSB repair and non-homologous end joining^{16,17}.

The mechanism by which translocations are formed is unknown, but could involve rejoining of broken chromosomes. To test this hypothesis, mutations in *LIG4* and *LIF1*, which encode ligase 4 (refs 22 and 23), were tested for their effect on the GCRs caused by *rad52*, *rfa1-t33*, *mre11*, *xrs2* and *rad27* mutations (Table 4). The *lig4* and *lif1* mutations only had small effects on the GCR rate, either as single mutations or in combination with *rad52*, *rfa1-t33*, *mre11*, *xrs2* and *rad27*. However, the *lig4* mutation decreased the translocation rate in *rad52*, *rfa1-t33* and *mre11* mutants by at least 5–10-fold. In contrast, the *lig4 rad27* double mutant had less than a twofold reduction in the rate of accumulating translocations compared with the *rad27* single mutant. This indicates that the bulk of the translocations that arise in this assay owing to defects in *RAD52*, *RFA1* or the *RAD50*–*MRE11*–*XRS2* complex require ligase 4—this could reflect either the involvement of non-homologous end joining, because the only known role of ligase 4 is in non-homologous end joining¹⁷, or a previously undetected function of ligase 4. *RAD50*, *MRE11* and *XRS2* are also required for non-homologous end joining¹⁷, so it is unexpected that *lig4* should reduce the rate of

Table 2 Interaction between checkpoint defects and telomerase defects

(a) GCR rate analysis

Relevant genotypes	<i>Pif1</i>		<i>pif1-m2</i>		<i>tlc1Δ</i>		<i>est2Δ</i>		<i>yku70Δ</i>	
	Strain number	Mutation rate (Can ⁺ 5-FOA)	Strain number	Mutation rate (Can ⁺ 5-FOA)	Strain number	Mutation rate (Can ⁺ 5-FOA)	Strain number	Mutation rate (Can ⁺ 5-FOA)	Strain number	Mutation rate (Can ⁺ 5-FOA)
Wild type	3615	3.5×10^{-10} (1)	4343	8.3×10^{-8} (237)	4224	3.1×10^{-10} (0.9)	4347	1.2×10^{-10} (0.3)	3696	1.2×10^{-10} (0.3)
<i>sm1Δ</i>	3733	3.1×10^{-10} (0.9)	4401	7.15×10^{-8} (204)		ND		ND		ND
<i>rfc5-1</i>	3727	6.6×10^{-8} (189)	4403	1.33×10^{-6} (3,800)	4411	5.0×10^{-8} (143)	4413	4.4×10^{-8} (125)	4417	2.5×10^{-8} (71)
<i>mec1Δ sm1Δ</i>	3735	6.8×10^{-8} (194)	4407	1.82×10^{-6} (5,200)	4130	6.3×10^{-8} (180)	4415	2.8×10^{-8} (80)	4419	4.3×10^{-8} (123)
<i>dun1Δ</i>	3739	7.3×10^{-8} (208)	4409	3.2×10^{-7} (914)		ND		ND		ND
<i>tel1Δ</i>	3731	2.0×10^{-10} (0.6)	4375	8.2×10^{-8} (234)	4501	1.0×10^{-6} (2,357)	4503	8.3×10^{-7} (2,372)		ND

(b) Rearrangement breakpoint analysis

Relevant genotypes	Wild type*		<i>pif1-m2</i>		<i>tlc1Δ</i>		Chromosome‡
	Telomere addition	Translocation	Telomere addition	Translocation	Telomere addition	Translocation	
Wild type	5	1,0	17	0,0	ND	ND	ND
<i>rfc5-1</i>	10	0,0	ND	ND	2†	1,5	2
<i>mec1Δ sm1Δ</i>	9	0,0	10	0,0	0	0,8	3
<i>tel1Δ</i>	0	0,6	8	0,2	0	2,4	4

Numbers in parentheses indicate the GCR rate relative to the wild-type GCR rate. The numbers reported in **b** are the actual numbers of each type of event seen. The translocation values in **b** indicate the number of non-homology breakpoints observed, followed by the number of micro-homology breakpoints observed. ND, not determined.

* Data from ref. 5.

† These two cases of telomere additions may represent translocations to telomeric regions.

‡ These events showed AC₁₋₃ repeats rather than TG₁₋₃ repeats typical of telomeres. They seem to represent cases where a broken chromosome V was fused to the end of another chromosome. Both *rfc5-1* *tlc1* cases showed the fusion of chromosome V to chromosome XII, although to different ends of chromosome XII. Two *mec1 sm1* *tlc1* cases showed the fusion of chromosome V to chromosome IV, and one *mec1 sm1* *tlc1* case showed the fusion of chromosome V to chromosome I. One *tel1* *tlc1* case showed the fusion of chromosome V to chromosome I. Breakpoint sequences from the other three *tel1* *tlc1* cases showed AC₁₋₃ repeats but we were unable to identify the sequences on the other side of the AC₁₋₃ repeats.

Table 3 Effect of mutations in recombination genes on the GCR rate

(a) GCR rate analysis				
Relevant genotypes	PIF1		pif1-m2	
	Strain number	Mutation rate (Can ⁺ 5-FOA)	Strain number	Mutation rate (Can ⁺ 5-FOA)
Wild type	3615	3.5×10^{-10} (1)	4343	8.3×10^{-8} (237)
<i>mre11Δ</i>	3633	2.2×10^{-7} (629)	4435	3.7×10^{-7} (1,045)
<i>xrs2Δ</i>	3634	1.9×10^{-7} (542)	4437	2.4×10^{-7} (685)
<i>rad50Δ</i>	3635	2.3×10^{-7} (657)		ND
<i>rad51Δ</i>	3636	3.5×10^{-9} (10)	4439	1.70×10^{-7} (486)
<i>rad52Δ</i>	4421	4.4×10^{-8} (126)	4441	1.1×10^{-8} (3,029)
<i>rad54Δ</i>	4473	1.9×10^{-9} (5)	4443	7.55×10^{-7} (2,157)
<i>rad57Δ</i> (30 °C)	3638	1.1×10^{-8} (31)	4445	6.3×10^{-7} (1,789)
<i>rad57Δ</i> (23 °C)	3638	4.1×10^{-8} (117)	4445	2.7×10^{-6} (7,714)
<i>rad59Δ</i>	4423	7.5×10^{-9} (21)	4447	2.3×10^{-7} (657)
<i>rdh54Δ</i>	4425	1.0×10^{-9} (3)	4449	4.7×10^{-8} (134)
<i>rad51Δ rad59Δ</i>	4427	1.2×10^{-7} (343)	4451	3.3×10^{-7} (943)
<i>rad51Δ mre11Δ</i>	3651	2.2×10^{-7} (629)		ND
<i>lig4Δ</i>	3641	1.6×10^{-9} (5)	4455	2.0×10^{-7} (569)
<i>rad51Δ lig4Δ</i>	4431	8.1×10^{-10} (2)	4457	1.1×10^{-7} (314)
<i>rad52Δ lig4Δ</i>	4433	8.8×10^{-8} (251)	4459	3.5×10^{-6} (10,000)

(b) Rearrangement breakpoint analysis

Relevant genotypes	PIF1		pif1-m2	
	Telomere addition	Translocation	Telomere addition	Translocation
Wild type	5*	1,0*	17	0,0
<i>mre11Δ</i>	3†	5,2†	14	0,0
<i>xrs2Δ</i>	4	2,3	12	0,0
<i>rad52Δ</i>	3	0,7	14	0,0†
<i>rdh54Δ</i>	ND	ND	10	0,0
<i>rad51Δ rad59Δ</i>	8	0,3	9	1,0
<i>lig4Δ</i>	6	0,0	9	0,0
<i>rad52Δ lig4Δ</i>	11	0,0	10	0,0

Numbers in parentheses indicate the GCR rate relative to the wild-type GCR rate. The numbers reported in **b** are the actual numbers of each type of event seen. The translocation values in **b** indicate the number of non-homology breakpoints observed, followed by the number of micro-homology breakpoints observed. ND, not determined.

* Data from ref. 5.

† Data from ref. 4.

‡ The sample size was too small to determine if translocations were eliminated in addition to the increase in the rate of telomere addition.

translocations caused by *mre11* or *xrs2* mutations. However, end joining can occur in *rad50*, *mre11*, *xrs2*, *yku70* and *yku80* mutants at high efficiency under some conditions, and the joints formed are different from the joints formed in wild-type cells^{17,24,25}. The

translocations that occur in a *rad27* mutant may be formed by a different mechanism because of the limited requirement for ligase 4 (ref. 4).

The data presented here support the following hypothesis for the suppression of GCRs (Fig. 1). DNA replication errors result in broken, stalled or collapsed replication forks that are processed to DSBs²⁶. These structures trigger S-phase checkpoint sensors, which results in phosphorylation of recombination and repair proteins, a transcriptional response and possibly cell-cycle arrest. The broken DNAs are normally repaired by some type of recombination. The genetic requirements for suppression of GCRs suggest that the most likely pathway involved is break-induced replication, and to a lesser extent DSB repair and non-homologous end joining. These substrates also yield telomere additions and translocations, which are seen as a result of mutations that inactivate recombination (see Table 3). Normally, *de novo* telomere additions are suppressed by PIF1, which suppresses GCRs (as shown here) because it is an inhibitor of telomerase activity that can suppress telomere additions at induced DSBs^{10,11}. By suppressing telomere additions, PIF1 channels substrates into non-mutagenic recombination. Thus, combinations of mutations that inactivate both pathways result in a synergistic increase in the GCR rate. Whether there is an active mechanism for suppressing translocations, in addition to recombination, is unclear. Mutations in S-phase checkpoint functions also increase the GCR rate, an effect attributed to defects in downstream phosphorylation and transcription events and possibly cell-cycle arrest defects⁵. S-phase checkpoint defects could also increase replication errors. Consequently, mutations that inactivate the S-phase checkpoint show synergistic interactions with mutations that inactivate recombination or suppression of telomere additions.

Genome instability is characteristic of cancer cells^{2,3,27}. The types of GCRs observed in the *S. cerevisiae* mutants studied here and elsewhere are similar to those seen in cancer cells^{2,3,27}. In many cases, including the RAD50–XRS2(NBS1)–MRE11(ATLD) complex, TEL1(ATM), MEC1(ATR), RAD53(CHK2), LIG4 and Ku70, defects in the mouse and human genes encoding homologues of these proteins are associated with genome instability and/or cancer susceptibility^{28,29}. The human homologue is shown in parentheses when the human and yeast names differ. Furthermore, the proteins encoded by the cancer susceptibility genes *BRCA1* and *BRCA2* either interact directly with proteins that function in the

Table 4 Ligase 4 is required for the formation of some translocations

(a) GCR rate analysis						
Relevant genotypes	Wild type		<i>lig4Δ</i>		<i>lif1Δ</i>	
	Strain number	Mutation rate (Can ⁺ 5-FOA)	Strain number	Mutation rate (Can ⁺ 5-FOA)	Strain number	Mutation rate (Can ⁺ 5-FOA)
Wild type	3615	3.5×10^{-10} (1)	3641	1.6×10^{-9} (5)	4465	4.0×10^{-10} (1.1)
<i>rad52Δ</i>	4421	4.4×10^{-8} (126)	4433	8.8×10^{-8} (251)		ND
<i>rfa1-t33</i>	3617	4.7×10^{-7} (1,342)	3649	2.9×10^{-7} (829)	4467	3.9×10^{-7} (1,114)
<i>mre11Δ</i>	3633	2.2×10^{-7} (629)	3652	1.7×10^{-7} (486)		ND
<i>rad27Δ</i>	3630	4.4×10^{-7} (1,257)	4463	4.1×10^{-7} (1,171)	4469	3.1×10^{-7} (886)
<i>xrs2Δ</i>	3634	1.9×10^{-7} (542)		ND	4471	2.8×10^{-7} (800)

(b) Rearrangement breakpoint analysis				
Relevant genotypes	LIG4		<i>lig4Δ</i>	
	Telomere addition	Translocation	Telomere addition	Translocation
Wild type	5*	1,0*	6	0,0
<i>rad52Δ</i>	3	0,7 [3.1×10^{-8}] (1)	11	0,0 [$<7.0 \times 10^{-9}$] (<0.2)
<i>rfa1-t33</i>	5†	0,6† [2.6×10^{-7}] (1)	10	0,2 [4.8×10^{-8}] (0.18)
<i>mre11Δ</i>	3†	5,2† [1.5×10^{-7}] (1)	9	0,0 [$<1.7 \times 10^{-8}$] (<0.1)
<i>rad27Δ</i>	4†	0,6† [2.6×10^{-7}] (1)	7	0,4 [1.5×10^{-7}] (0.58)

Numbers in parentheses in **a** indicate the GCR rate relative to the wild-type GCR rate. The numbers reported in **b** are the actual numbers of each type of event seen. Translocation values in **b** indicate the number of non-homology breakpoints observed, followed by the number of micro-homology breakpoints observed. ND, not determined. Square brackets in **b** indicate the rate of translocations, and round parentheses indicate the rate of translocations in the *lig4Δ* derivative relative to the *LIG4* strain.

* Data from ref. 5.

† Data from ref. 4.

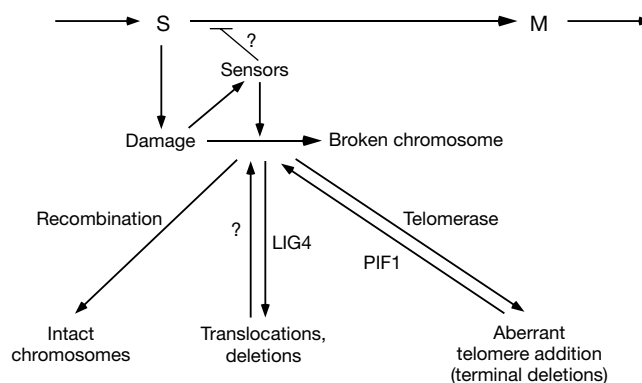


Figure 1 Pathways for suppression of genome instability. Integration of replication damage and the cell cycle with pathways that suppress genome instability. DNA replication damage activates S-phase checkpoint sensors that activate downstream responses, including repair, required for suppression of genome instability. Cell-cycle arrest (indicated by a question mark) may not be necessary for suppression of genome instability (see ref. 5). The replication damage is acted on by multiple pathways; the

primary non-mutagenic repair pathway seems to be break-induced replication although DSB repair and non-homologous end joining may have a small involvement. The principal mutagenic pathways are the translocation pathway(s) and the *de novo* telomere-addition pathway suppressed by PIF1. How many translocation pathways exist and whether there is specific suppression of translocations is not known.

genome instability suppression pathways described here or are phosphorylated by proteins that function in these pathways²⁹. Although many functions have been attributed to these cancer susceptibility genes, our results suggest that they could all function directly in the suppression of genome instability by acting in the types of pathways described here. □

Methods

Strains and gene disruption of *S. cerevisiae*

Media for propagation of strains were as previously described^{4,5}. All yeast strains were propagated at 30 °C except strains containing the *rfc5-1* mutation, which were grown at 23 °C. Strains containing mutations in genes encoding telomerase components (*tlc1*, *est1*, *est2*, *est3*) were propagated for no more than 50 generations before expansion of cultures for fluctuation analysis. All strains were derived from the S288c parental strain RDY3023 (*MATa*, *ura3-52*, *leu2Δ1*, *trp1Δ63*, *his3Δ200*, *lys2ΔBgl*, *hom3-10*, *ade2Δ1*, *ade8*) and in addition contained the *hxt13::URA3* insertion used in the GCR assay^{4,5}. Gene disruptions were made by standard PCR-based gene disruption methods and verified by PCR as described⁵. The sequences of primers used to generate disruption cassettes and to confirm disruptions are available on request. Strains carrying mutations of *pif1-m1*, *pif1-m2*, or *cdc13-2* were generated using pop-in and pop-out plasmids and confirmed by analysing the phenotypes of the mutant strains and by sequencing the region of the gene containing the mutation as described^{10,30}. Strains with the *stn1-13* mutation were generated as described⁸.

Gross chromosome rearrangements

All GCR rates were determined by fluctuation analysis two or more times using either 5 or 11 cultures, and the average value is reported^{4,5}. The sequences of independent rearrangement breakpoints were determined and classified as previously published^{4,5}.

Received 7 March; accepted 3 May 2001.

- Mitelman, F. *Catalog of Chromosome Aberration in Cancer* (Wiley Liss, New York, 1991).
- Lengauer, C., Kinzler, K. W. & Vogelstein, B. Genetic instabilities in human cancers. *Nature* **396**, 643–649 (1998).
- Padilla-Nash, H. M. *et al.* Molecular cytogenetic analysis of the bladder carcinoma cell line BK-10 by spectral karyotyping. *Genes Chromosom. Cancer* **25**, 53–59 (1999).
- Chen, C. & Kolodner, R. D. Gross chromosomal rearrangements in *Saccharomyces cerevisiae* replication and recombination defective mutants. *Nature Genet.* **23**, 81–85 (1999).
- Myung, K., Datta, A. & Kolodner, R. D. Suppression of spontaneous chromosomal rearrangements by the S-phase checkpoint in *Saccharomyces cerevisiae*. *Cell* **104**, 397–408 (2001).
- Lundblad, V. DNA ends: maintenance of chromosome termini versus repair of double strand breaks. *Mut. Res.* **451**, 227–240 (2000).
- Bryan, T. M. & Cech, T. R. Telomerase and the maintenance of chromosome ends. *Curr. Opin. Cell Biol.* **11**, 318–324 (1999).
- Grandin, N., Reed, S. I. & Charbonneau, M. Stn1, a new *Saccharomyces cerevisiae* protein, is implicated in telomere size regulation in association with Cdc13. *Genes Dev.* **11**, 512–527 (1997).
- Lahaye, A., Stahl, H., Thines-Sempoux, D. & Foury, F. PIF1: a DNA helicase in yeast mitochondria. *EMBO J.* **10**, 997–1007 (1991).
- Schulz, V. & Zakian, V. A. The *Saccharomyces* PIF1 DNA helicase inhibits telomere elongation and *de novo* telomere formation. *Cell* **76**, 145–155 (1994).
- Zhou, J.-Q., Monson, E. K., Teng, S.-C., Schulz, V. P. & Zakian, V. A. Pif1p helicase, a catalytic inhibitor of telomerase in yeast. *Science* **289**, 771–774 (2000).

- Diede, S. J. & Gottschling, D. E. Telomerase-mediated telomere addition *in vivo* requires DNA primase and DNA polymerases α and δ . *Cell* **99**, 723–733 (1999).
- Chen, C., Umez, K. & Kolodner, R. D. Chromosomal rearrangements occur in *S. cerevisiae* *rfa1* mutator mutants due to mutagenic lesions processed by double-strand-break repair. *Mol. Cell* **2**, 9–22 (1998).
- Lee, S. E., Paques, F., Sylvan, J. & Haber, J. E. Role of yeast *SIR* genes and mating type in directing DNA double-strand breaks to homologous and non-homologous repair paths. *Curr. Biol.* **9**, 767–770 (1999).
- Pennock, E., Buckley, K. & Lundblad, V. Cdc13 delivers separate complexes to the telomere for end protection and replication. *Cell* **104**, 387–396 (2001).
- Haber, J. E. Exploring the pathways of homologous recombination. *Curr. Opin. Cell Biol.* **4**, 401–412 (1992).
- Lewis, L. K. & Resnick, M. A. Tying up loose ends: nonhomologous end-joining in *Saccharomyces cerevisiae*. *Mut. Res.* **451**, 71–89 (2000).
- Lee, S. E. *et al.* *Saccharomyces* Ku70, Mre11/Rad50, and RPA proteins regulate adaptation to G2/M arrest after DNA damage. *Cell* **94**, 399–409 (1998).
- Signon, L., Malkova, A., Naylor, M. L., Klein, H. & Haber, J. E. Genetic requirements for *RAD51*- and *RAD54*-independent break-induced replication repair of a chromosomal double-strand break. *Mol. Cell Biol.* **21**, 2048–2056 (2001).
- Bai, Y. & Symington, L. S. A Rad52 homolog is required for *RAD51*-independent mitotic recombination in *Saccharomyces cerevisiae*. *Genes Dev.* **10**, 2025–2037 (1996).
- Chen, Q., Ijima, A. & Greider, C. W. Two survivor pathways that allow growth in the absence of telomerase are generated by distinct telomere recombination events. *Mol. Cell Biol.* **21**, 1819–1827 (2001).
- Teo, S. H. & Jackson, S. P. Identification of *Saccharomyces cerevisiae* DNA ligase IV: involvement in DNA double-strand break-repair. *EMBO J.* **16**, 4788–4795 (1997).
- Herrmann, G., Lindahl, T. & Schar, P. *Saccharomyces cerevisiae* LIF1: a function involved in DNA double-strand break repair related to mammalian XRCC4. *EMBO J.* **17**, 4188–4198 (1998).
- Moore, J. K. & Haber, J. E. Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* **16**, 2164–2173 (1996).
- Boulton, S. J. & Jackson, S. P. *Saccharomyces cerevisiae* Ku70 potentiates illegitimate DNA double-strand break repair and serves as a barrier to error-prone DNA repair pathways. *EMBO J.* **15**, 5093–5103 (1996).
- Michel, B. Replication fork arrest and DNA recombination. *Trends Biochem. Sci.* **25**, 173–178 (2000).
- Vessey, C. J., Norbury, C. J. & Hickson, I. D. Genetic disorders associated with cancer predisposition and genomic instability. *Prog. Nucleic Acid Res. Mol. Biol.* **63**, 189–221 (1999).
- Petrini, J. H. J. The Mre11 complex and ATM: collaborating to navigate S phase. *Curr. Opin. Cell Biol.* **12**, 293–296 (2000).
- Khanna, K. K. & Jackson, S. P. DNA double-strand breaks: signaling, repair and the cancer connection. *Nature Genet.* **27**, 247–254 (2001).
- Nugent, C. L., Hughes, T. R., Lue, N. F. & Lundblad, V. Cdc13p: A single-strand telomeric DNA-binding protein with a dual role in yeast telomere maintenance. *Science* **274**, 249–252 (1996).

Acknowledgements

We would like to thank V. Zakian, S. Lee, J. Haber, V. Lundblad, L. Symington and M. Charbonneau for plasmids and helpful discussions; P. Lau and K. Schmidt for comments on the manuscript; and J. Weger and J. Green for DNA sequencing. This work was supported by a NIH grant to R.D.K. and a Damon Runyon-Walter Winchell Foundation Postdoctoral Fellowship to K.M.

Correspondence and requests for materials should be addressed to R.D.K. (e-mail: rkolodner@ucsd.edu).

Crystal structure of a complex of a type IA DNA topoisomerase with a single-stranded DNA molecule

Anita Changela*, Russell J. DiGate† & Alfonso Mondragón*

* Department of Biochemistry, Molecular Biology and Cell Biology, Northwestern University, Evanston, Illinois 60208, USA

† Department of Pharmaceutical Sciences, University of Maryland School of Pharmacy, Baltimore, Maryland 21201, USA

A variety of cellular processes, including DNA replication, transcription, and chromosome condensation, require enzymes that can regulate the ensuing topological changes occurring in DNA. Such enzymes—DNA topoisomerases—alter DNA topology by catalysing the cleavage of single-stranded DNA (ssDNA) or double-stranded DNA (dsDNA), the passage of DNA through the resulting break, and the rejoining of the broken phosphodiester backbone¹. DNA topoisomerase III from *Escherichia coli* belongs to the type IA family of DNA topoisomerases, which transiently cleave ssDNA via formation of a covalent 5' phosphotyrosine intermediate. Here we report the crystal structure, at 2.05 Å resolution, of an inactive mutant of *E. coli* DNA topoisomerase III in a non-covalent complex with an 8-base ssDNA molecule. The enzyme undergoes a conformational change that allows the oligonucleotide to bind within a groove leading to the active site. We note that the ssDNA molecule adopts a conformation like that of B-DNA while bound to the enzyme. The position of the DNA within the realigned active site provides insight into the role of several highly conserved residues during catalysis. These findings confirm various aspects of the type IA topoisomerase mechanism while suggesting functional implications for other topoisomerases and proteins that perform DNA rearrangements.

We formed an *E. coli* DNA topoisomerase III–ssDNA complex by using an inactive mutant of the enzyme in which the catalytic tyrosine (Tyr328) was replaced with phenylalanine (TopoIII-Y328F). This mutant displays the same substrate binding preferences as the wild-type enzyme². Topoisomerase III recognizes preferentially a specific cleavage sequence consisting of seven nucleotides, with at least one nucleotide on the 3' side of the cleavage site². Co-crystals were obtained with intact TopoIII-Y328F in the presence of an 8-base oligonucleotide containing a preferred cleavage sequence with one nucleotide on the 3' side of the putative

cleavage site (5'-CGCAACT↓T-3'). The affinity of TopoIII-Y328F for this oligonucleotide was determined by fluorescence polarization to be ~400–500 nM (see Methods). We solved the structure of the complex by molecular replacement using the apo topoisomerase III structure as a search model³. Excellent electron density was seen for the entire length of the oligonucleotide, which allowed unambiguous oligonucleotide sequence assignment and determination of DNA chain polarity. The model was refined to 2.05 Å resolution with a final *R* factor of 23% and an *R*_{free} of 26% with excellent stereochemistry (see Table 1 in Supplementary Information).

While complexed with ssDNA, topoisomerase III maintains the characteristic toroidal shape of type IA topoisomerases, although a rearrangement of its domains relative to the apo enzyme is observed. The ssDNA is bound within a groove extending from the active site at the interface of domains I, III and IV, and continuing through a region created primarily by domains I and IV (Fig. 1). This groove has been noted previously as a putative ssDNA binding region³. The carboxy terminus of the protein, which has been implicated in DNA binding⁴, extends from domain IV. Although the last 33 residues of the protein are disordered in the complex structure, there is interpretable electron density for 11 residues at the C terminus that were disordered in the apo enzyme structure. Although these last residues do not interact with the oligonucleotide in the complex structure, their path suggests that the missing C-terminal domain travels back towards the main body of the protein where it could interact with the passing DNA strand during the reaction cycle. Topoisomerase III contains a positively charged loop (decatenation loop) in domain IV that extends from the central hole, and which has been shown to be critical for efficient decatenation activity by the enzyme⁵. In the complex, binding of ssDNA results in a slight displacement of the decatenation loop with respect to other regions in domain IV, yet the loop makes no contacts to the oligonucleotide.

The mode of ssDNA binding that we observe is representative of the pre-cleavage or post-religation state of the enzyme in association with the scissile DNA strand during catalysis. In the complex, the putative scissile bond is identified by the orientation of Phe 328 towards the phosphate group between Cyt 6 and Thy 7. Although this cleavage site does not correspond to the major one predicted from the sequence, it is consistent with a minor cleavage site². The oligonucleotide is probably bound in a manner that maximizes protein–DNA contacts at the expense of specificity. The ssDNA exhibits a 3' to 5' polarity from the active site towards the C terminus, and is positioned along the groove such that the phosphate groups are buried within the enzyme while the bases either point towards the bottom of the groove or face outwards. Although

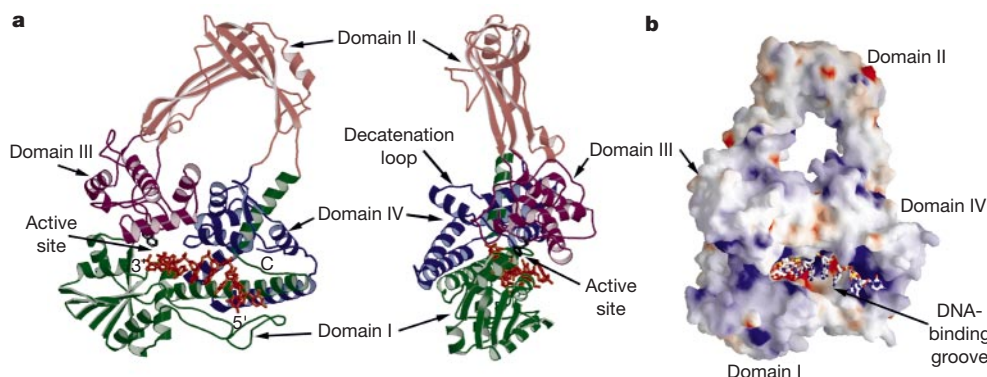


Figure 1 Overall structure of the *E. coli* DNA topoisomerase III–ssDNA complex. **a**, Ribbon representations of the front and side views of the complex. Domains I–IV of the enzyme are coloured green, pink, purple and blue, respectively. The ssDNA molecule is shown in red. The active-site Tyr 328, which was mutated to a phenylalanine, is shown in black ball-and-stick. **b**, Surface representation of *E. coli* DNA topoisomerase III, revealing the

prominent groove where ssDNA binds. The surface is coloured to represent electrostatic potential ($-15k_B T/e$ to $+15k_B T/e$) where $k_B T$ is the product of the Boltzmann constant and temperature and e is the electron charge, with negatively and positively charged regions shown in red and blue, respectively. The DNA is represented as sticks (carbon, yellow; nitrogen, blue; oxygen, red; and phosphorous, yellow).

this orientation leaves several bases exposed for possible base pairing, the oligonucleotide binds deeply within the groove in such a way that dsDNA would be sterically hindered from binding in the groove.

The conformation adopted by ssDNA is remarkably similar to that of one strand of a B-DNA molecule (Fig. 2a). The similarity is not confined to the phosphodiester backbone, as the six bases at the 3' end form stacking interactions and closely follow the conformation of B-DNA. Type IA topoisomerases only cleave ssDNA regions, so it is surprising that the conformation of the oligonucleotide resembles that of B-DNA, especially near the active site. Modelling of B-DNA into the ssDNA-binding cleft of topoisomerase III reveals that although the binding groove is complementary to ssDNA, it does not allow dsDNA to bind (Fig. 2b). Nevertheless, duplex DNA could be accommodated beyond the active site of the enzyme upon a significant movement of domain III. The specificity of the binding groove for ssDNA probably guides the recognition of a ssDNA region as required for cleavage by type IA topoisomerases. Moreover, by stabilizing its scissile ssDNA substrate in a B-DNA-like conformation near the active site, the enzyme could facilitate the subsequent transition back to duplex DNA.

The enzyme interacts with all eight nucleotides of its DNA substrate through direct hydrogen bonds, electrostatic interactions, and water-mediated contacts (Supplementary Information). Most of these interactions are made to phosphate groups. Whereas many of the residues contacting the phosphodiester backbone are highly

conserved among type IA topoisomerases, most of the residues involved in base contacts are not and thus are unlikely to function in base-specific recognition. After cleavage, the directionality of the ssDNA observed in the structure and the extensive protein–DNA contacts within the binding groove enable the enzyme to sequester the 3'-OH end of the cleaved DNA strand within the DNA binding groove while the 5' phosphoryl end remains covalently attached to the active-site tyrosine in domain III. Non-covalent association of the broken strand within the binding groove effectively anchors it in place for religation and prevents its interference during the strand passage events.

Comparison of the structure of topoisomerase III in the presence and absence of ssDNA reveals that the enzyme undergoes a domain movement upon binding ssDNA. Superposition of domain I shows how domains II and III, along with part of domain IV, swing away from the rest of the protein by $\sim 20^\circ$ when the enzyme is bound to ssDNA (Fig. 3a). This conformational change enables the enzyme to interact with the entire length of the oligonucleotide, and also provides access to the active site. Individual superposition of each domain using C α atoms gives root-mean-square deviation (r.m.s.d.) values for domains I, II, III and IV of 0.86 Å, 0.66 Å, 0.48 Å and 3.05 Å, respectively, thus indicating that only domain IV undergoes any significant internal rearrangement. In order to accommodate the DNA, the last two α -helices that lead into the C terminus shift with respect to the rest of domain IV. Most of the other elements in domain IV associate largely with domains II and

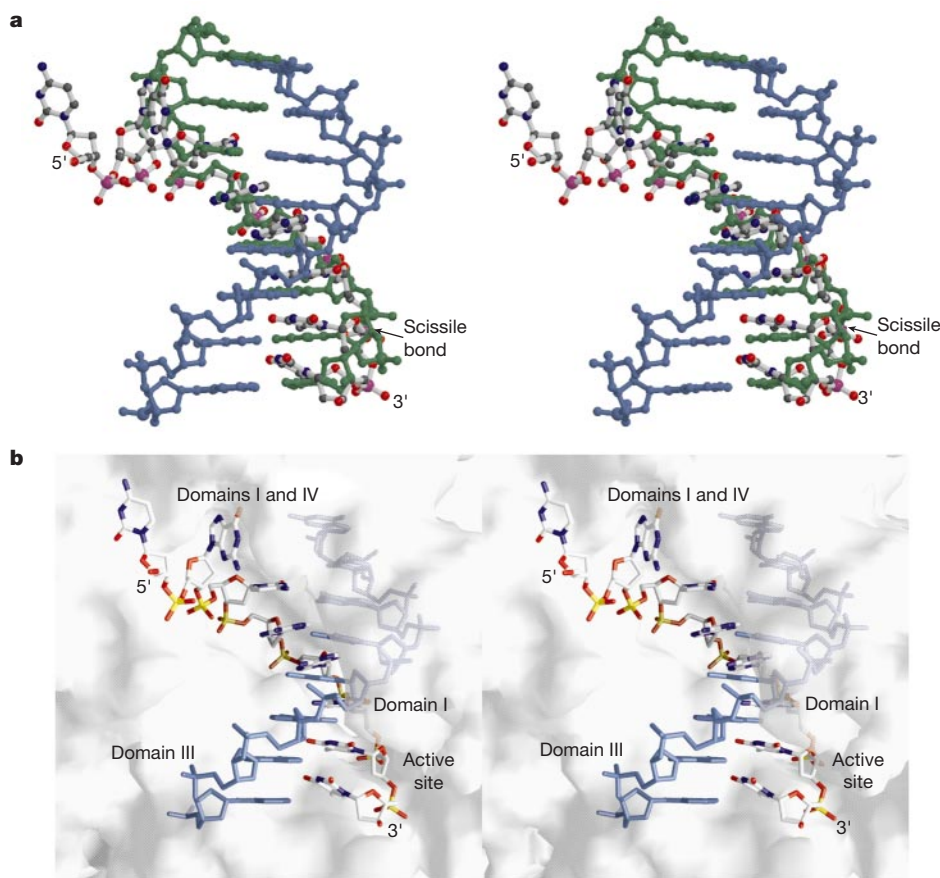


Figure 2 Comparison of the bound ssDNA structure to B-DNA. **a**, Stereo view showing the superposition of the bound ssDNA molecule onto one strand of an ideal B-DNA model. Superposition of the two molecules using the six nucleotides at the 3' end of the oligonucleotide gives an r.m.s.d. value of 1.7 Å for all atoms. The bound ssDNA is depicted with grey bonds and coloured by atom, while the complementary strands of the B-DNA model are shown in green and blue. **b**, Stereo view of the modelling of B-DNA into

the ssDNA binding groove. A hybrid B-DNA molecule was generated with the bound ssDNA molecule (coloured by atom) and a complementary strand from the B-DNA model (shown in blue). The faded regions of the complementary strand represent areas of the modelled DNA strand that would collide with the protein and prevent binding within the cleft.

III during the conformational change.

The apparent hinge element contributing to the overall conformational change resides within domain I at a region where the enzyme contacts the DNA (Fig. 3b, c). The amino-terminal end of the helix leading from domain I into domain II points directly into the phosphodiester backbone of the ssDNA. Residues 197–199, located at the N terminus of this helix, interact with several phosphate groups near the 5' end of the oligonucleotide. Orientation of the helix dipole into the negatively charged sugar-phosphate backbone probably stabilizes enzyme–substrate interactions in this region. Comparison of the apo and DNA-bound forms of the enzyme reveals that this helix has shifted slightly with respect to the rest of domain I. This small movement leads into the more pronounced displacement of domains II and III, thus suggesting that protein–DNA interactions at the base of the shifted helix may be responsible for triggering the overall conformational change.

The movement of domains II and III results in the creation of an active site different from that observed in the apo forms of type IA topoisomerases^{3,6} (Fig. 4a). The phenyl ring of Phe 328 is oriented directly towards the phosphodiester backbone at a distance of 3.6 Å; a tyrosine in this position would be well placed for nucleophilic attack on the phosphodiester bond. The phosphate group targeted for nucleophilic attack is coordinated via its non-bridging oxygens to Arg 330 and Lys 8. Arg 330, invariant among type IA topoisomerases, interacts directly with the phosphate groups of Thy 7 and Thy 8, thereby securing the 3' end of the oligonucleotide within the active site during cleavage and to domain III after cleavage. Comparison of Lys 13 in *E. coli* topoisomerase I and Lys 8 in topoisomerase III shows that, in both cases, the lysine is in a position to interact with the phosphate group of the scissile bond, thus suggesting a conserved functional role for this residue. Lys 8 also forms a salt bridge with the highly conserved Asp 103, which, in turn, participates in a charged network with Glu 7 and His 381, both invariant residues (Fig. 4b). Glu 7 directly contacts the oligonucleotide at the bridging 3'-oxygen atom of the putative scissile bond.

The realignment of conserved active-site residues from domains I and III provides insight into the cleavage mechanism used by type IA topoisomerases. The proposed mechanism (Fig. 4c) shares similarities to reactions performed by other proteins catalysing phosphoryl transfer through a tyrosine nucleophile using acid/base catalysis⁷. The manner in which the catalytic tyrosine is activated for nucleophilic attack still remains unclear, as the structure of the complex presents no obvious candidates for a general base that could abstract the proton from the tyrosine hydroxyl. By analogy to type IB topoisomerases, which function through a 3' phosphotyrosine linkage, one possibility is that a nearby water molecule⁸ facilitates abstraction of the proton from the tyrosine. During cleavage, Arg 330 and Lys 8 are positioned to stabilize the negative charge developing on the non-bridging oxygens of the pentavalent transition state. The structure reveals that Glu 7 is ideally placed to donate a proton for the stabilization of the leaving 3'-oxygen upon cleavage of the scissile strand. If religation occurs by the microscopic reversal of cleavage, Glu 7 is also in place to abstract a proton from the free 3'-hydroxyl group thereby activating it for nucleophilic attack on the phosphotyrosyl intermediate. A role for Glu 7 in catalysis is further supported by mutagenesis studies, which have shown that this residue is essential for activity^{9,10}.

Magnesium is required by type IA topoisomerases for the relaxation of negatively supercoiled DNA^{11,12}. Although magnesium stimulates DNA cleavage, its presence is not required for cleavage¹² and its role in religation remains unclear. Asp 103, Asp 105 and Glu 107 are conserved acidic residues from domain I that have been suggested to be involved in magnesium binding^{6,13}. In the complex, neither Asp 105 nor Glu 107 makes any contacts to the bound oligonucleotide, to Asp 103, or to each other. The configuration

of these acidic residues in a pre-cleavage state with ssDNA presents an unlikely metal binding site. Among DNA topoisomerases, magnesium has been observed only in the structure of the archaeal topoisomerase VI (ref. 14). In an attempt to find a possible magnesium binding site in the presence of DNA, we soaked cocrystals of the *E. coli* DNA topoisomerase III–ssDNA complex with a magnesium-containing solution (see Methods), yet there was no evidence of a magnesium ion in the resulting structure. Under these experimental conditions, the conserved acidic residues do not bind magnesium although it is still possible that they coordinate metal during a stage in the reaction other than cleavage. The structure suggests that magnesium may stimulate the overall relaxation reaction before or after cleavage, perhaps in DNA binding or release, but it does not directly participate in DNA cleavage.

Despite lack of sequence similarity, structural similarities between type IA and type II topoisomerases, which cleave dsDNA, have led to the proposal of a unified mechanism for these two families of enzymes¹⁵. In addition to the common use of a phosphotyrosine intermediate, a feature linking type IA and type II families is an α/β domain or Rossmann fold classified as a TOPRIM domain due to its presence in several topoisomerases and primases¹⁶. This α/β domain, corresponding to part of domain I in type IA topoisomerases, was assigned a putative role in ssDNA binding when first observed in the structure of *E. coli* DNA topoisomerase I (ref. 6). The functional importance of the TOPRIM fold has been extended to include a role in catalysis by virtue of a highly conserved glutamate residue (Glu 7 in topoisomerase III) shown to be required for activity. The structure of *E. coli* primase^{17,18} shows the presence of a TOPRIM fold, thus reinforcing a link to the DNA topoisomerases. Although the functional relevance of this connection is not yet established, analogy to the structure of

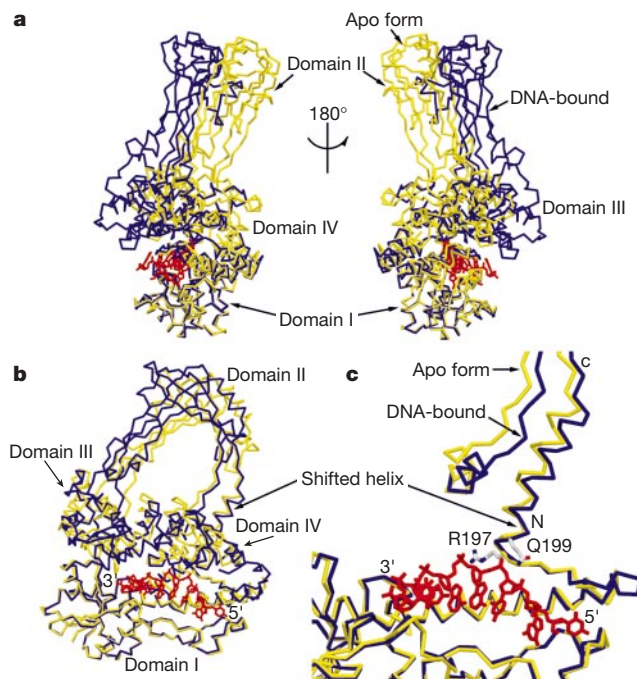


Figure 3 Conformational changes on ssDNA binding. **a**, Side views of the superposition of *E. coli* DNA topoisomerase III in the presence and absence of DNA. Residues 1–198 in domain I of the apo enzyme (yellow) and the DNA-bound enzyme (blue) were superimposed. The ssDNA molecule is shown in red. The orientation on the left shows domain IV facing out, while on the right, the molecules have been rotated 180° such that domain III is facing out. **b**, Front view of superposition. **c**, Close-up view of a putative hinge point. The shifted helix (residues 198–215) leads from domain I into domain II. At the base of the helix, Arg 197 and the invariant Gln 199 are shown interacting with the phosphodiester backbone.

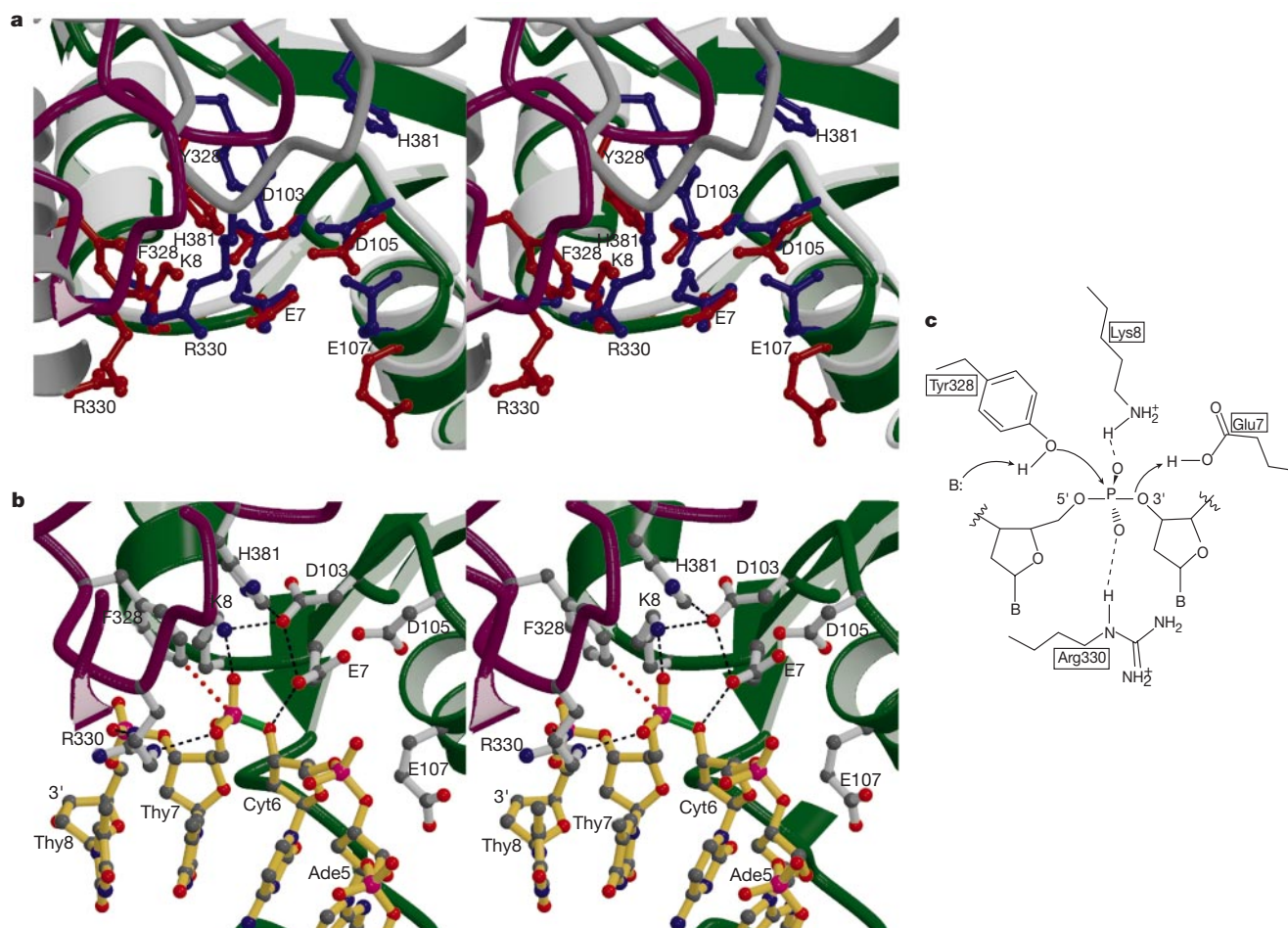


Figure 4 Active site and catalytic mechanism. **a**, Stereo view showing the *E. coli* DNA topoisomerase III active site in the presence and absence of DNA. Residues from the apo structure are shown in blue, while residues of the enzyme–DNA complex are coloured red. Secondary structural elements of domains I and III of the apo enzyme are depicted in light and dark shades of grey, respectively. Elements corresponding to domains I and III of the complex are coloured green and purple, respectively. **b**, A stereo view showing the active-site region formed at the interface of domains I and III in the presence of ssDNA. The bonds in the oligonucleotide are coloured yellow, and protein elements are coloured by domain as in Fig. 1a. Hydrogen bonding and electrostatic interactions are depicted by

black dotted lines. A red dotted line highlights the orientation of Phe 328 towards the phosphate group of the putative scissile bond (green). **c**, Schematic diagram illustrating a possible mechanism of DNA cleavage used by type IA topoisomerases. A nearby group acting as a general base (B:) abstracts the proton from the tyrosine thus activating it for nucleophilic attack. In *E. coli* DNA topoisomerase III, Tyr 328 forms a covalent linkage with the 5' phosphate of the scissile bond while the pentavalent transition state is stabilized by Arg 330 and Lys 8. Cleavage of the ssDNA is facilitated by Glu 7 which acts as a general acid catalyst by donating a proton to the leaving 3'-oxygen atom of the scissile bond.

the complex supports a putative role of the TOPRIM domain in bacterial primases in nucleic-acid binding and catalysis of the phosphotransfer reaction.

In type IA and type II topoisomerases, the TOPRIM fold and the catalytic tyrosine are located in different domains. The structure of the topoisomerase III–ssDNA complex shows how DNA binding triggers a relatively small conformational change that realigns the TOPRIM domain with the catalytic tyrosine in domain III, thus resulting in the formation of a functional active site. It is plausible that DNA binding in type II enzymes also results in a conformational change that brings the two distal catalytic domains together. Biochemical characterization of yeast topoisomerase II suggests type II enzymes act in *trans* whereby the catalytic tyrosine from one monomer associates with the acidic residues of the TOPRIM fold of the second monomer¹⁹. Two structures of yeast topoisomerase II, representing different conformations adopted by the homodimeric enzyme, suggest that conformational changes of the magnitude required for creation of an active site involving the TOPRIM domain are possible^{20,21}. This idea is supported by the structure of another type II topoisomerase, the archaeal topoisomerase VI, in which the tyrosine residue from one monomer lies

approximately 10 Å from the TOPRIM fold of the other¹⁴. In this case, a slight reorientation of domains, similar to what is observed in the topoisomerase III–ssDNA structure, would be sufficient to align the two domains for catalysis.

The structure of *E. coli* DNA topoisomerase III in a non-covalent complex with ssDNA represents, to our knowledge, the first detailed picture of how type IA topoisomerases interact with their scissile DNA substrate. These findings help further our understanding of a general mechanism by which a protein can catalyse nucleic-acid rearrangements through complex domain movements triggered by substrate binding. □

Methods

Crystallization and data collection

A mutated gene encoding a catalytically inactive form of *E. coli* DNA topoisomerase III (TopoIII-Y328F) was cloned into the pET21 vector. The protein was expressed with a C-terminal six-histidine tag in BL21 cells by induction with bacteriophage λ CE6 at 37 °C. After cell lysis, full-length TopoIII-Y328F was purified using standard nickel affinity chromatography followed by gel filtration (Sephacryl S-100) chromatography and dialysed into 50 mM Tris (pH 8.0), 200 mM NaCl, 1 mM dithiothreitol, and 1 mM EDTA for crystallization. Oligonucleotides were commercially synthesized (Integrated DNA Technologies), purified by HPLC, and concentrated to 15–30 mg ml⁻¹ in water for

crystallization. The sequence of the oligonucleotide that yielded cocrystals is 5'-CGCAACT↓T-3' where ↓ denotes the preferred site of cleavage. The binding affinity of TopoIII-Y328F for this oligonucleotide was measured by fluorescence polarization with 5'-fluorescein-labelled DNA. Reactions were prepared in 50 mM Tris (pH 8.0) and 100 mM NaCl, and allowed to equilibrate for 3 min at 25 °C before taking polarization measurements.

The complex was prepared for crystallization by incubating TopoIII-Y328F (4 mg ml⁻¹) with a twofold molar excess of oligonucleotide at 4 °C for 30 min. Cocrystals formed at 22 °C in hanging drops equilibrated against 1.3 M (NH₄)₂SO₄, 0.1 M sodium citrate (pH 5.5), and 0.8 M NaCl. Crystals appeared in 2–3 d and grew as thin plates to maximum dimensions of ~0.05 × 0.2 × 1 mm³ within 2 weeks. The crystals belong to space group C2 with cell dimensions of *a* = 122.0 Å, *b* = 60.8 Å, *c* = 125.4 Å, and β = 90.7°, and contain one complex per asymmetric unit. Before data collection, the crystals were quickly transferred in a single step to mother liquor supplemented with 25% glycerol and flash-cooled in liquid nitrogen. For the magnesium studies, crystals were soaked for 16 h in mother liquor containing 10 mM MgCl₂ before transfer to the cryoprotectant containing 10 mM MgCl₂. Macromolecular crystal annealing was performed to decrease crystal mosaicity²². Crystals diffracted to 2.0 Å at 100 K with synchrotron radiation, although they still displayed a large mosaicity (>1.5°). Diffraction data were collected in 0.5° oscillation steps due to the large mosaicity, processed with the program XDS²³, and scaled with SCALA²⁴. Data collection statistics are available (see Table I in Supplementary Information).

Structure determination and refinement

The structure of the complex was solved by molecular replacement with AMORE²⁴ using diffraction data from a single crystal. The initial search model consisted of domains I and IV (residues 1–216; 489–609) from the 3.0 Å resolution structure of *E. coli* DNA topoisomerase III⁷. Rotation and translation searches performed with data to 3 Å yielded a clear solution for this region of the protein. Domains I and IV were then fixed, and a solution for domains II and III together was found that gave an *R*-factor of 49.8% for the entire model. Initial refinement was carried out to 3 Å, and consisted of rigid body refinement of the individual domains interspersed with cycles of simulated annealing and conjugate gradient minimization in CNS²⁵. The resolution was gradually extended to 2.05 Å as the *R*-factor decreased. Initial electron-density maps were improved by density modification in CNS, and showed unambiguous density for all eight bases of the oligonucleotide. Building of the DNA model and rebuilding of the protein model was carried out in program O³⁵, and all refinement was done in CNS. A large region of domain IV had undergone rearrangements and needed to be rebuilt. Density could be seen for a short surface loop (185–190) and additional C-terminal residues that were disordered in the apo topoisomerase III structure. A strong difference density peak (>8σ) present near the 5' end of the oligonucleotide was assigned as a chloride ion. The structure was refined to an *R* factor of 23.3% and an *R*_{free} of 26.0%, with 90.5% of all residues in the most favoured regions of the Ramachandran plot and no residues in the disallowed regions. The final model contains residues 1–620 of TopoIII-Y328F, the entire oligonucleotide, 332 water molecules, one sulphate molecule, and one chloride ion. Refinement statistics are available (see Table I in Supplementary Information). Figures were made with the programs GRASP²⁷, MOLSCRIPT²⁸, RASTER3D²⁹ and SETOR³⁰.

Received 4 January; accepted 2 May 2001.

1. Wang, J. C. DNA topoisomerases. *Annu. Rev. Biochem.* **65**, 635–692 (1996).
2. Zhang, H. L., Malpure, S. & DiGate, R. J. *Escherichia coli* DNA topoisomerase III is a site-specific DNA binding protein that binds asymmetrically to its cleavage site. *J. Biol. Chem.* **270**, 23700–23705 (1995).
3. Mondragon, A. & DiGate, R. The structure of *Escherichia coli* DNA topoisomerase III. *Struct. Fold Des.* **7**, 1373–1383 (1999).
4. Zhang, H. L. & DiGate, R. J. The carboxyl-terminal residues of *Escherichia coli* DNA topoisomerase III are involved in substrate binding. *J. Biol. Chem.* **269**, 9052–9059 (1994).
5. Li, Z., Mondragon, A., Hiasa, H., Mariani, K. J. & DiGate, R. J. Identification of a unique domain essential for *Escherichia coli* DNA topoisomerase III-catalysed decatenation of replication intermediates. *Mol. Microbiol.* **35**, 888–895 (2000).
6. Lima, C. D., Wang, J. C. & Mondragon, A. Three-dimensional structure of the 67K N-terminal fragment of *E. coli* DNA topoisomerase I. *Nature* **367**, 138–146 (1994).
7. Sherratt, D. J. & Wigley, D. B. Conserved themes but novel activities in recombinases and topoisomerases. *Cell* **93**, 149–152 (1998).
8. Redinbo, M. R., Champoux, J. J. & Hol, W. G. Novel insights into catalytic mechanism from a crystal structure of human topoisomerase I in complex with DNA. *Biochemistry* **39**, 6832–6840 (2000).
9. Chen, S. J. & Wang, J. C. Identification of active site residues in *Escherichia coli* DNA topoisomerase I. *J. Biol. Chem.* **273**, 6050–6056 (1998).
10. Zhu, C. X., Roche, C. J., Papanicolaou, N., DiPietrantonio, A. & Tse-Dinh, Y. C. Site-directed mutagenesis of conserved aspartates, glutamates and arginines in the active site region of *Escherichia coli* DNA topoisomerase I. *J. Biol. Chem.* **273**, 8783–8789 (1998).
11. Wang, J. C. Interaction between DNA and an *Escherichia coli* protein ω. *J. Mol. Biol.* **55**, 523–533 (1971).
12. Domanico, P. L. & Tse-Dinh, Y. C. Mechanistic studies on *Escherichia coli* DNA topoisomerase I: Divalent ion effects. *J. Inorg. Biochem.* **42**, 87–96 (1991).
13. Zhu, C. X. & Tse-Dinh, Y. C. The acidic triad conserved in type IA DNA topoisomerases is required for binding of Mg(II) and subsequent conformational change. *J. Biol. Chem.* **275**, 5318–5322 (2000).
14. Nichols, M. D., DeAngelis, K., Keck, J. L. & Berger, J. M. Structure and function of an archaeal topoisomerase VI subunit with homology to the meiotic recombination factor Spo11. *EMBO J.* **18**, 6177–6188 (1999).
15. Berger, J. M., Fass, D., Wang, J. C. & Harrison, S. C. Structural similarities between topoisomerases that cleave one or both DNA strands. *Proc. Natl Acad. Sci. USA* **95**, 7876–7881 (1998).
16. Aravind, L., Leipe, D. D. & Koonin, E. V. Toprim—a conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins. *Nucleic Acids Res.* **26**,

4205–4213 (1998).

17. Keck, J. L., Roche, D. D., Lynch, A. S. & Berger, J. M. Structure of the RNA polymerase domain of *E. coli* primase. *Science* **287**, 2482–2486 (2000).
18. Podobnik, M., McInerney, P., O'Donnell, M. & Kuriyan, J. A TOPRIM domain in the crystal structure of the catalytic core of *Escherichia coli* primase confirms a structural link to DNA topoisomerases. *J. Mol. Biol.* **300**, 353–362 (2000).
19. Liu, Q. & Wang, J. C. Similarity in the catalysis of DNA breakage and rejoining by type IA and IIA DNA topoisomerases. *Proc. Natl Acad. Sci. USA* **96**, 881–886 (1999).
20. Berger, J. M., Gamblin, S. J., Harrison, S. C. & Wang, J. C. Structure and mechanism of DNA topoisomerase II. *Nature* **379**, 255–232 (1996).
21. Fass, D., Bogden, C. E. & Berger, J. M. Quaternary changes in topoisomerase II may direct orthogonal movement of two DNA strands. *Nature Struct. Biol.* **6**, 322–326 (1999).
22. Harp, J. M., Timm, D. E. & Bunick, G. J. Macromolecular crystal annealing: overcoming increased mosaicity associated with cryocrystallography. *Acta Crystallogr. D* **54**, 622–628 (1998).
23. Kabsch, W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26**, 795–800 (1993).
24. Collaborative Computational Project 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
25. Brunger, A. T. et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
26. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
27. Nicholls, A., Sharp, K. A. & Honig, B. H. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Genet.* **11**, 281–286 (1991).
28. Kraulis, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
29. Merritt, E. A. & Murphy, M. E. P. Raster 3D version 2.0. A program for photorealistic molecular graphics. *Acta Crystallogr. D* **50**, 869–873 (1994).
30. Evans, S. V. SETOR: Hardward lighted three-dimensional solid model representations of macromolecules. *J. Mol. Graphics* **11**, 134–138 (1993).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements

We acknowledge the contributions of H. Feinberg to the early stages of this project, and thank W. Niu for technical assistance. We thank K. Perry, E. Sontheimer and J. Widom for comments and suggestions. Research was supported by the NIH (A.M.) and an NRSA Institutional Training Grant in Molecular Biophysics (A.C.). We acknowledge the use of instruments in the Keck Biophysics Facility at Northwestern University. Portions of this work were performed at the DuPont-Northwestern-Dow collaborative Access Team (DND-CAT) Synchrotron Research Center at the Advanced Photon Source and at the Stanford Synchrotron Radiation Laboratory (SSRL). DND-CAT is supported by DuPont, Dow and the NSF, and use of the APS is supported by the DOE. SSRL is operated by the DOE, Office of Basic Energy Sciences. The SSRL Biotechnology Program is supported by the NIH and the DOE.

Correspondence and requests for materials should be addressed to A.M. (e-mail: a-mondragon@northwestern.edu). Coordinates have been deposited in the Protein Data Bank under the accession code 1I7D.

correction

A conserved XIAP-interaction motif in caspase-9 and Smac/DIABLO regulates caspase activity and apoptosis

Srinivasa M. Srinivasula, Ramesh Hegde, Ayman Saleh, Pinaki Datta, Eric Shiozaki, Jijie Chai, Ryung-Ah Lee, Paul D. Robbins, Teresa Fernandes-Alnemri, Yigong Shi & Emad S. Alnemri

Nature **410**, 112–116 (2001).

It has been brought to our attention by Paul Robbins (University of Pittsburgh) that the wrong western blot may have been used by Ayman Saleh in his laboratory to generate the Casp-9 (WT) control panel (second panel from the top) of Fig. 2b. However, we have repeated the experiment and the results and conclusions are unchanged. □

Contacts

Publisher: Fabien Savenay

Editor: Paul Smaglik

Sales Director: Ben Crowe

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: classified@nature.com

Group European Manager:

Leonie Welss (4954)

European Manager:

Nevin Bayoumi (4978)

European Sales Consultant:

Jonathan Bull (4973)

UK/ RoW/ Ireland:

Matt Powell (4953), Ben Corp (4974),

Andy Douglas (4975)

Holland/ Italy: Nevin Bayoumi (4978)

Scandinavia: Sille Opstrup (4994)

Production Manager:

Billie Franklin
To send films and materials use
London address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: classified@nature.com

International

Advertising Coordinator:

Laura Pearson (4977)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Offices

France/ Belgium:

Christine Niox-Chateau

Tel + 33 (0) 1 43 20 16 51

Fax + 33 (0) 1 43 20 51 52

e-mail: c.nioxchateau@nature.com

Spain/ Portugal:

Amanda Wren

Tel + 34 91 3913835

Fax + 34 91 319 0091

e-mail: a.wren@nature.com

Germany/ Austria/ Switzerland:

Patrick Phelan/ Kate Turner

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mails: p.phelan@nature.com

k.turner@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707

Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: classified@nature.com

US Sales Director:

Ben Crowe

US Sales Manager:

Peyton Mason

US Advertising Coordinator:

Corissa Salzman

Japan Head Office, Tokyo

Shin-Mitsuke Building (4F),
3-6 Ichigaya Tamachi Shinjuku-Ku,
Tokyo 162

Tel +81 3 3267 8751

Fax +81 3 3267 8746

e-mail: k.cowan@nature.com

Japan Manager:

Kate Cowan

Tracking tenure

Academia in general — and science in particular — is a meritocracy, with the spoils of tenure and higher salaries going to the best and brightest. But a recent report on salary and tenure in the United States shows that the number of people in the higher ranks is shrinking. And the difference in the way the best and the rest are rewarded is becoming more pronounced.

According to a recent report by the US National Education Association based on US Department of Education data for 1998–99 (<http://www.nea.org/he/heupdate/vol7no3.pdf>), tenured professors at private research institutions earned \$31,000 more on average than their non-tenure-track colleagues, and tenured faculty at public research institutions earned \$26,000 more than their colleagues at the same universities. Meanwhile, the percentage of all faculty with tenure fell from 35% in the 1992–93 Department of Education survey to 32% in this year's report.

The analysis does not draw a perfect picture of science salaries because the report does not break down the data by academic disciplines. But departments in science and technology tend to be among the highest paid, so it seems safe to derive some conclusions from the report. It also meshes broadly with other reports, such as the US National Science Foundation's science and engineering indicators, and a 2000–01 study by the American Association of University Professors.

If this trend does indeed apply to science, it could bode ill for the long-term health of academia. Universities should not be surprised if more people leave for industrial careers, or do not even consider academia. A secure position in academia was once considered the height of meritocracy. But, with the odds of advancement tightening, it will not be surprising if the best and the brightest pursue other options.

Paul Smaglik

Naturejobs editor



Contents

MOVERS

Fun at Genentech, NASA high-fliers and more [Back page](#)

WWW.NATURE.COM/ NATUREJOBS

Career centre
Information on the scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS

